The Library of
St. Olaf College
Northfield, Minn.

Accession No. 30548

Class 517.8 Vol.

FEB 26 1930

Digitized by the Internet Archive
in 2024

https://archive.org/details/functionsforreal0000unse

American Mathematical Series
E. J. TOWNSEND
GENERAL EDITOR

## American Mathematical Series

Under the Editorship of

E. J. TOWNSEND

**College Algebra.** By H. L. Rietz and A. R. Crathorne

**Introductory College Algebra.** By H. L. Rietz and A. R. Crathorne

**School Algebra.** First Course. By H. L. Rietz, A. R. Crathorne and E. H. Taylor

**School Algebra.** Second Course. By H. L. Rietz, A. R. Crathorne and E. H. Taylor

**Mathematics of Finance.** By H. L. Rietz, A. R. Crathorne and J. C. Rietz

**Analytic Geometry.** By L. W. Dowling and F. E. Turneaure

**Analytic Geometry of Space.** By Virgil Snyder and C. H. Sisam

**Plane Geometry.** By J. W. Young and A. J. Schwartz

**Functions of a Complex Variable.** By E. J. Townsend

**Functions of Real Variables.** By E. J. Townsend

**Essentials of Calculus.** By E. J. Townsend and G. A. Goodenough

**Analytic Geometry.** By Walter B. Ford

**Essentials of Trigonometry.** By A. R. Crathorne and E. B. Lytle

# FUNCTIONS
# OF
# REAL VARIABLES

BY

E. J. TOWNSEND, Ph.D.
PROFESSOR OF MATHEMATICS, UNIVERSITY OF ILLINOIS

NEW YORK
HENRY HOLT AND COMPANY

QA331
.5
.T74

COPYRIGHT, 1928
BY
HENRY HOLT AND COMPANY

PRINTED IN THE U.S.A.

# PREFACE

The present volume is based upon a course in Real Variables given for some years at the University of Illinois and is intended as a companion volume to the author's book on Complex Variables. It presupposes such knowledge of mathematics as the student will have had who has completed the usual undergraduate course in mathematics including a year's course in calculus and perhaps an elementary course in differential equations or a course in advanced calculus. The purpose of the course is to discuss those topics which will enable the student to have a better grasp and understanding of the fundamental principles of the calculus of real variables and know something of the more recent developments of this branch of analysis.

As an essential tool in the development of the subject, there has been introduced some of the more elementary notions of point sets. For convenience in reference, or to the reader who wishes to become familiar with this subject only, the general discussion of point sets is presented in a single chapter. In using the book as the basis of a course in Real Variables, it may be found more convenient to postpone certain portions of this chapter until they are actually needed in the discussions which follow. Thus, the articles on content and measure may well be postponed until the subject of integration is considered.

It has been the purpose to limit the amount of material to what can readily be covered in a year's course of three or four hours per week, and to select those topics for discussion which the student might well know who expects to teach calculus or to continue his study of analysis. In accordance with this plan, there will be found from time to time a general statement of facts, made without formal proof, where these facts are not essential to the logical presentation of an introductory course in the subject. In such cases, reference is made to the original sources of information for the use of the reader who wishes to verify the conclusions stated. These references may well be made topics of special report in the class room by those students who desire to cover the subject more fully. The dates of some of the more important theorems, as well as other historical data, are given in foot-notes for the convenience of those who wish to trace the historical development of the subject.

The book was first used in mimeograph form as a text in the course given by the author. Many suggestions of those taking the course have been incorporated, which it is hoped have added to the value of the book as a class room text. The author wishes to acknowledge his obligations to Professor W. B. Ford of the University of Michigan and Dr. C. M. Huber of Rutgers University who have read the galley proof sheets. He wishes also to acknowledge his indebtedness to Professor T. H. Hildebandt of the University of Michigan who has offered many valuable suggestions for the improvement of the chapters on integration. While these friends are not in any way responsible for such defects as the book may possess, their encouragement and criticisms have been most helpful in its preparation.

E. J. T.

University of Illinois
*August*, 1928

# CONTENTS

## CHAPTER I

## REAL NUMBER SYSTEM

ARTICLE PAGE

**1.** Rational Numbers ........ 1
**2.** Sequences of Rational Numbers ........ 3
**3.** Existence of Irrational Numbers ........ 7
**4.** Cantor's Theory of Irrational Numbers ........ 9
**5.** Dedekind's Theory of Irrational Numbers ........ 11
**6.** The Real Number System ........ 14
**7.** Geometric Representation of Real Numbers ........ 17
**8.** Base of a System of Notation ........ 22

## CHAPTER II

## THEORY OF POINT SETS

**9.** Infinite Aggregates ........ 25
**10.** Transfinite Numbers ........ 26
**11.** Fundamental Definitions ........ 28
**12.** Fundamental Operations with Point Sets ........ 32
**13.** Sequence of Point Sets ........ 33
**14.** Properties of Derived Sets ........ 35
**15.** Properties of Enumerable Sets ........ 37
**16.** Open and Closed Sets ........ 44
**17.** Perfect Sets ........ 46
**18.** Point Sets of the First and Second Category ........ 50
**19.** Non-Linear Sets ........ 56
**20.** Content of Point Sets ........ 58
**21.** Measure of Point Sets ........ 64

## CHAPTER III

## CONTINUITY AND DISCONTINUITY OF FUNCTIONS

**22.** Fundamental Definitions ........ 79
**23.** Definition of a Limit ........ 82
**24.** Laws of Operation with Limits ........ 88
**25.** Continuity and Discontinuity of a Function ........ 98

ARTICLE PAGE
26. Properties of Continuous Functions 107
27. Semi-Continuous Functions 123
28. Discontinuous Functions 127

Chapter IV

## DERIVATIVES AND THEIR PROPERTIES

29. Definition of a Derivative and of a Derivative Number 149
30. Continuous Functions Having no Derivative 155
31. Properties of Derivatives and Derivative Numbers 160
32. Definition of Partial Derivatives 170
33. Properties of Partial Derivatives 171
34. Total Differentiability 177
35. Interchange of the Order of Differentiation 189

Chapter V

## RIEMANN THEORY OF INTEGRATION

36. Definition of a Riemann Integral 198
37. Upper and Lower Integrals 200
38. Functions of Limited Variation 202
39. Existence Theorems for Riemann Integrals 209
40. Integrals of Bounded Functions 215
41. Indefinite Integrals 225
42. Definition of Infinite Integrals 234
43. Convergence of Integrals with Infinite Limits of Integration 241
44. Convergence of Integrals of Non-Bounded Functions 245
45. Properties of Infinite Integrals 248
46. Integrals Involving a Parameter 250
47. Uniform Convergence of an Infinite Integral 264
48. Double Integrals. Repeated Integrals 268

Chapter VI

## LEBESGUE AND OTHER INTEGRALS

49. Measurable Functions 277
50. Definition of a Lebesgue Integral 285
51. Lebesgue Integrals of Bounded Functions 289
52. Comparison of Riemann and Lebesgue Integrals 295
53. Lebesgue Integrals for Non-Bounded Functions 297
54. Indefinite Lebesgue Integrals 306
55. Lebesgue Integrals of Functions of Two Independent Variables 323
56. Other Definitions of Integration 327

## CHAPTER VII

## INFINITE SERIES

ARTICLE PAGE
**57.** General Definitions........ 333
**58.** Laws of Operation with Series........ 338
**59.** Uniform Convergence of a Series........ 345
**60.** Simple-Uniform Convergence. Quasi-Uniform Convergence........ 353
**61.** Integration of Series........ 358
**62.** Differentiation of Series........ 363
**63.** Power Series........ 367
**64.** Condensation of Singularities........ 375
**65.** Divergent Series........ 387
INDEX........ 403

# ERRATA

Page 4. Line 1 below (2), read $n \geqq m$ for $n > m$.

“ 24. Ex. 12, read $\left\{\left(1+\frac{1}{n}\right)^n\right\}$ for $\left\{\left(1+\frac{1}{n}\right)\right\}^n$.

“ 45. Line 19, read “within” for “with.”

“ 47. Line 2, Theorem IV, insert “everywhere dense, non-abutting and” after “set of.”

“ 52. Line 12 from bottom, read $k \leqq m+n$ for $k = m+n$.

“ 60. Line 4 below Theorem I, read “finite” for “arbitrarily small.”

“ 64. Theorem V, line 1, read “a bounded sequence of discrete sets.”

“ 71. Line 4, read $m[C(A)]+m(E_2)$.

“ 84. Line 1 below Fig. 12, read $\underset{x\to 0}{L}$ for $\underset{n\to\infty}{L}$.

“ 87. Ex. 5, line 1, read $y^6$ for $y^3$ in denominator of the fraction.

“ 91. Lines 3, 4 should read: $A-\epsilon_1 \leqq f(x) \leqq A+\epsilon_2$, or $-\epsilon_1 \leqq f(x)-A \leqq \epsilon_2$.

“ 92. Line 5 of Ex. 3 read $0\cdot y$ in numerator of fraction for $0\cdot x$.

“ 95. Line 7 from bottom, read $|A-f(\overline{x})|$ for $|A-f(x)|$.

“ 97. Lines 16, 20, 23, read $2\eta_k$ for $\eta_k$.

“ 127. Line 5 from bottom, read “small, positive number” for “selected.”

“ 129. Line 3, Theorem III, read “small, positive” for “chosen.”

“ 143. Line 3, read $\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x,y)=f(x_0,y_0)$.

“ 145. Line 22, read $A \geqq f_n(x) \geqq B$.

“ 146. Line 3, omit “or less.”

“ 150. Line 6 from bottom, read “the oscillation of $F(x_0, \Delta x)$” for “if $f'(x_0)$ exists, the saltus of $f'(x)$ at $x_0$”; in (2) read “$\omega(F, x_0)$” for “$\omega(f', x_0)$.”

“ 176. Line 11, read $R_1$ for $R$.

“ 209. Inequality (1) should read $\underline{S}(\sigma) \leqq S(\sigma) \leqq \overline{S}(\sigma)$.

“ 212. Cor. line 3, read “small, positive number” for “chosen member.”

“ 241. Insert $\frac{1}{2}$ before the integral $\int_0^\infty \frac{\sin t}{\sqrt{t}}\,dt$.

“ 244. Line 2 below Theorem V, read $\leqq$ for $<$ between the integrals given; Ex. 5, read $\int_1^\infty$ for $\int^\infty$.

“ 284. Line 4 from bottom, read XI for IX.

“ 293. Line 11, read $A\sum_{k=1}^{n} m(E_k)$ for $Am(E)$.

“ 305. Line 4 from bottom, read “for every value of $\epsilon > 0$ and” for “we select an $\epsilon$ such that.”

“ 306. Line 1 from bottom, omit “of.”

“ 322. Line 18, read “the function $f'(x)$” for “the function $f(x)$.”

“ 357. Line 2 from bottom, omit “be continuous”; line 3 from bottom, insert “may be” after “function.”

“ 366. Line 2 below Theorem III, read $\int_a^x S'_n(x)dx$.

“ 367. Line 4, read “Theorem IX” for “Theorem III.”

“ 398. Line 1 from bottom, insert $=$ after $u(x, y)$.

Note: The omission of a limit of integration from integral signs has not been listed where it can be supplied without ambiguity.

NORTHFIELD, MINN.
30548

# FUNCTIONS OF REAL VARIABLES

## CHAPTER I

## REAL NUMBER SYSTEM

**1. Rational numbers.** The process of counting involves a knowledge of the positive integers 1, 2, 3, . . . , often called the **natural numbers.** One also obtains this class of numbers by assuming the concept unity and the knowledge of a certain process called addition. When we admit the inverse operation of subtraction, we are led to introduce negative integers and zero in order always to interpret the results.

Again, when we divide one integer by another, we find it necessary to introduce another class of numbers in order always to be able to interpret the results, namely **fractions.**

The totality of numbers thus far considered, namely positive and negative integers, zero, and fractions constitute a class of numbers called the **rational numbers.** These numbers constitute a number-system having certain properties; for example, they constitute a closed set with respect to the processes of addition, subtraction, multiplication, and division. In other words, if these operations are performed upon rational numbers, the results are always rational numbers. Every rational number can be expressed as the quotient of one integer by another; that is, it can be expressed in the form $\frac{p}{q}$, where $p$ and $q$ are integers prime to each other and $q$ is different from zero.

Division by zero must be excluded from consideration, since it can not lead to a definite result; for, we should then have a rational number $k$ such that

$$\frac{p}{0} = k,$$

and, therefore, if we admit the fundamental laws of arithmetic,

$$p = 0 \cdot k.$$

This relation allows $k$ to be any rational number, which is contrary to the assumption that it was a particular number.

By the process of addition and subtraction of unity, the integral numbers may be placed in a definite order of rank, so that we are able to say that one integer is less than or greater than another; that is, one has a higher or a lower rank than the other. Of two rational numbers, we say that the one is **less than** or **greater than** the other if, when reduced to a common denominator, the one numerator is less than or greater than the other. When arranged in order of rank, an aggregate of numbers is said to be **ordered.** If $x_1$, $x_2$ are the two numbers, we express symbolically the fact that $x_1$ is less than or greater than $x_2$ by writing respectively

$$x_1 < x_2, \; x_1 > x_2.$$

The number $x$ is said to lie **between** $x_1$ and $x_2$ if we have

$$x_1 < x < x_2,$$

where $x_1$ is less than $x_2$. Between any two rational numbers, there is always another rational number. To show this, all we need to do is to take one-half of the sum of the two given rational numbers. The resulting number is rational and lies between the two given numbers. This process of subdivision may be carried on indefinitely, and hence the foregoing statement is equivalent to saying that between any two rational numbers there exist an infinite number of other rational numbers.

We shall denote the **absolute value** of a number $x_1$ by writing $| x_1 |$, read "the absolute value of $x_1$", by which we mean its numerical value independently of whether it is positive or negative. Thus, we have $| x_2 | = | -x_2 |$.

We shall speak of a number as being **arbitrarily small** if it may be chosen as small as we please but is different from zero. We shall commonly denote such a number by $\epsilon$. As applied to positive numbers, the expression "as small as we please" is understood to refer to the rank of the number in an ordered system.

The elements of one aggregate are said to be in a **one-to-one correspondence** with those of another when to each element in the one there corresponds one and only one element of the other, and conversely. Thus, there exists a one-to-one correspondence between the chairs and the persons in a room when to each person may be assigned one and only one chair, and conversely with every chair there may be associated one and only one person. Likewise, there is a one-to-one correspondence between the terms of an infinite series and the positive integers, since we can put the subscripts 1, 2, 3, . . . on the terms.

**2. Sequences of rational numbers.** An ordered, infinite aggregate of rational numbers

$$a_1, a_2, \cdots \quad a_n, \cdots$$

having a one-to-one correspondence with the positive integers in their natural order is called an infinite sequence, or more briefly a **sequence of rational numbers.** The individual numbers composing the sequence are called the **elements** of the sequence. A sequence always has a first element, and to each element there is a definite one which follows. Moreover, each element, except the first, is preceded by a finite number of other elements and every one is followed by an infinite number of elements. The sequence

$$a_1, a_2, \cdots, a_n, \cdots$$

will be denoted symbolically by $\{a_n\}$. We shall now consider some of the special properties of sequences of rational numbers as a basis for a more precise consideration of certain fundamental definitions and operations.

It is to be noted that a set of numbers may form a sequence if arranged in one way, but not if arranged in some other manner. Thus, the set of natural numbers form a sequence if arranged in the order of rank. However, if we take first the succession of even integers followed by the succession of odd integers, we should not have a sequence; for, such an arrangement does not have a one-to-one correspondence with the positive integers which preserves the order of arrangement of the numbers.

If the absolute values of the elements of a sequence do not exceed a definite value, then the sequence is called a **bounded sequence.**

A sequence of rational numbers $\{a_n\}$ is said to **converge** if, for each arbitrarily small positive rational value assigned to $\epsilon$, there exists a definite integer $m$, such that for every integer $n > m$, we have

$$| a_n - a_{n+p} | < \epsilon, p = 1, 2, 3, \cdots.$$

Thus, if there exists a rational number $A$ such that for all the values of $n$ greater than a certain integer $m$, we have

$$| A - a_n | < \epsilon,$$

where $\epsilon$ is an arbitrarily small positive rational number, the sequence $\{a_n\}$ is said to have the rational limit $A$, and we write

$$\underset{n \to \infty}{L} \{a_n\} = A.$$

If A is not rational, the limiting process has no meaning in a number system composed only of rational numbers. A convergent sequence is also called a **regular sequence.**

We have the following theorems concerning regular sequences.

THEOREM I. *Every regular sequence of rational numbers is bounded.*

Let the given sequence of rational numbers be

$$\{a_n\} \equiv a_1, a_2, \cdots, a_n, \cdots.$$

We must show that there exists a finite positive integer $M$ such that

$$|a_n| < M, n = 1, 2, 3, \cdots.$$

Since the sequence is convergent, it follows from the foregoing definition that there exists an integer $m$ such that for every $n > m$ we have

$$|a_{n+p} - a_n| < \epsilon, p = 1, 2, 3, \cdots. \tag{1}$$

Let $M$ be equal to or greater than the numerical value of each of the numbers

$$a_1, a_2, \cdots, a_m, a_m - \epsilon, a_m + \epsilon. \tag{2}$$

From (1) it follows that for $n > m$, we have

$$a_m - \epsilon < a_n < a_m + \epsilon;$$

that is, from $a_m$ on, all subsequent elements must differ from $a_m$ by less than $\epsilon$. Since $M$ is equal to or greater than the absolute value of every number given in (2), it follows that

$$|a_n| < M,$$

whatever value is assigned to $n$.

It must not be inferred from this theorem that a convergent sequence is always increasing or always decreasing. It may change from an increasing to a decreasing sequence, or *vice versa*, but in no case can an element exceed in absolute value a certain number. There are sequences which never decrease and others which never increase. Such sequences are said to be **monotone.** A sequence which never decreases is called a monotone increasing sequence, and one which never increases is called a monotone decreasing sequence. Thus

$$1, 1 + \frac{1}{1\cdot2}, 1 + \frac{1}{1\cdot2} + \frac{1}{2\cdot3}, 1 + \frac{1}{1\cdot2} + \frac{1}{2\cdot3} + \frac{1}{3\cdot4}, \cdots$$

is a monotone increasing sequence;

$$1, 1 - \tfrac{1}{2}, 1 - \tfrac{1}{2} - \tfrac{1}{3}, 1 - \tfrac{1}{2} - \tfrac{1}{3} - \tfrac{1}{4}, \cdots$$

is a monotone decreasing sequence. The sequence,

$$1 - \frac{1}{3!},\ 1 - \frac{1}{3!} + \frac{1}{5!},\ 1 - \frac{1}{3!} + \frac{1}{5!} - \frac{1}{7!},\ \cdots,$$

on the other hand, is not monotone.

THEOREM II. *If a regular sequence of rational numbers does not converge to the limit zero, its elements must be from some point on either all positive or all negative.*

Let us suppose the contrary to be true; namely, let us assume that there is no point beyond which the elements are either all positive or all negative. Then, however large we select $n$, there exists a value of $p$ such that $a_{n-p}$ and $a_n$ have opposite signs. Since the given sequence $\{a_n\}$ is regular, we have

$$| a_{n+p} - a_n | < \epsilon,\ p = 1, 2, 3, \cdots, n > m. \qquad (3)$$

For those values of $p$ for which $a_{n-p}$ and $a_n$ have opposite signs, the absolute value of the difference between them must be greater than either number. Consequently, the only way that the relation (3) can exist is that we have

$$| a_{n+p} | < \epsilon,\ | a_n | < \epsilon.$$

However small we select the number $\epsilon$, it follows that there exists some element of the sequence such that it and all following elements are numerically less than $\epsilon$. Consequently, the sequence must converge to the limiting value zero, but this conclusion is contrary to the assumed hypothesis. Hence, the theorem follows as stated.

THEOREM III. *A bounded monotone sequence of rational numbers is convergent.*

We shall assume that the given sequence does not converge to zero; for, in that case the theorem is established. As the sequence is monotone, it follows that the elements are ultimately all positive or all negative. We shall prove the theorem for monotone increasing sequences whose terms are ultimately all positive. The proofs for decreasing sequences and for sequences whose terms are ultimately all negative are similar and are left to the reader to supply.

Let $\{a_n\}$ be the given sequence; then by hypothesis we have for $n$ greater than a definite $m$

$$| a_n | < M,$$

where $M$ is some finite positive integer.

We shall prove the theorem by assuming that the given sequence is not convergent and showing that this assumption leads to a contradiction of the given hypothesis. We assume then that a positive integer $m$ does not exist such that for $n > m$ we have

$$| a_{n+p} - a_n | < \epsilon, \ p = 1, 2, 3, \cdots ; \tag{4}$$

that is, however large we select $m$ there exists an $n > m$ such that for some positive integer $p$ we have

$$| a_{n+p} - a_n | \geqq \epsilon.$$

Denote by $n_k$ and $p_k$ the corresponding values of $n$ and $p$ for which this condition is true; that is, the relation (4) is false. Since the given sequence is monotone increasing and its elements ultimately all positive, we have

$$\begin{aligned} &a_{n_1+p_1} - a_{n_1} \geqq \epsilon, \\ &a_{n_2+p_2} - a_{n_2} \geqq \epsilon, \text{ where } n_2 = n_1 + p_1, \\ &\cdots\cdots\cdots\cdots \\ &a_{n_k+p_k} - a_{n_k} \geqq \epsilon, \text{ where } n_k = n_{k-1} + p_{k-1}, \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

Making use of the relations between $n_k$ and $n_{k-1} + p_{k-1}$, we have

$$\begin{aligned} a_{n_2} &= a_{n_1+p_1} \geqq a_{n_1} + \epsilon, \\ a_{n_3} &= a_{n_2+p_2} \geqq a_{n_1} + 2\epsilon, \\ &\cdots\cdots\cdots \\ a_{n_{k+1}} &= a_{n_k+p_k} \geqq a_{n_1} + k\epsilon, \\ &\cdots\cdots\cdots \end{aligned}$$

For a given $\epsilon$, $k$ may be taken sufficiently large so that the product $k \cdot \epsilon$ may be made as large as we please; for, it will be seen that $\epsilon$ is first chosen and that $k$ may be taken as large as we please. Hence, $k \cdot \epsilon$ may be made larger than $M$. The rational numbers $a_{n_k}$ are elements of the given sequence. The conclusion that $k \cdot \epsilon$ may be made to exceed $M$, and hence $k$ may be so chosen that

$$a_{n_{k+1}} \geqq a_{n_1} + k\epsilon > M,$$

contradicts the hypothesis that the given sequence is bounded. From this contradiction the theorem follows for the case considered; namely, where the elements of the sequence are all positive and the sequence is monotone increasing.

We assume with Cantor that corresponding to every regular sequence of rational numbers there exists a uniquely determined object called a **real number.** We express this fact by saying that every regular sequence of rational numbers defines a real number, which is represented by the sequence. Every rational number itself can be represented by such a regular sequence. Thus, the rational number $a$ is represented by the sequence

$$a.\ a,\ a,\ \cdots,$$

and since the real number uniquely defined by this sequence corresponds to the rational number $a$, it serves the purpose of analysis to say that it is $a$.

While every regular sequence of rational numbers defines one and only one real number, it does not follow that a given real number may not be represented by more than one such sequence.

Thus, the sequence

$$1,\ \tfrac{3}{2},\ \tfrac{2}{3},\ \tfrac{5}{4},\ \tfrac{4}{5},\ \tfrac{7}{6},\ \tfrac{6}{7},\ \cdots$$

is regular since the difference between the $n^{th}$ element and any succeeding one may be made smaller than any arbitrarily small positive number $\epsilon$ by taking $n$ sufficiently large. This sequence defines the number 1, as we may see by inspection. In the same way it may be shown that the number 1 is also defined by either of the following regular sequences

$$1,\ 1,\ 1,\ 1,\ \cdots,$$
$$\tfrac{1}{2},\ \tfrac{2}{3},\ \tfrac{3}{4},\ \tfrac{4}{5},\ \tfrac{5}{6},\ \cdots.$$

**3. Irrational numbers.** While the class of real numbers includes all rational numbers, we shall now see that it also includes some that are not rational. Indeed, the system of rational numbers is not adequate for the purposes of analysis. For, when the operation of root extraction is included in the processes of arithmetic, we can not always interpret our results in terms of rational numbers. For example, $\sqrt{2}$ can not be so interpreted; that is, it can not be expressed in the form $\frac{p}{q}$, where $p$ and $q$ are integers prime to each other. To show that this is the case, let us put

$$\sqrt{2} = \frac{p}{q},$$

and examine the consequences of such an assumption. Squaring both sides, we have

$$2 = \frac{p^2}{q^2}, \text{ or } p^2 = 2q^2.$$

The number 2 must then be a factor of $p^2$, and since $p^2$ is a square, each of its factors must appear an even number of times. Hence, 2 must appear an even number of times on the one side of this equation and consequently the same number of times on the other; that is, it must appear an even number of times as a factor in $2q^2$. Consequently, 2 is a factor of $q^2$, and again as this is a square it must appear an even number of times in $q^2$. It follows then from the given assumption that 2 appears as an even number of times as a factor of one member of this equation and as an odd number of times of the other member. This is impossible and hence our assumption that $\sqrt{2}$ is a rational number is false.

However, the process of extracting the square root of 2 gives as the successive approximations the sequence

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ \cdots.$$

The difference between the $n^{th}$ element and any succeeding one can not exceed $\frac{1}{10^{n-1}}\cdot$ By taking $n$ sufficiently large, we can take this difference as small as we please, and consequently the definition of a convergent sequence of rational numbers is satisfied and, in accordance with the assumption that every such sequence defines a real number, we may conclude that this sequence defines in the limit the number $\sqrt{2}$.

It is evident from the foregoing discussion that the class of real numbers includes some numbers like $\sqrt{2}$ which are not rational. We shall call those real numbers which are not rational **irrational numbers.**[1] The ratio of the diameter of a circle to its circumference, the base of the system of natural logarithms, etc., furnish other illustrations of irrational numbers already familiar to the reader.

Whenever we extend our number system by the addition of a new class of numbers, it becomes necessary to define for the extended class of numbers the fundamental operations of arithmetic, and this must be done in a manner consistent with the definitions of these operations for the more restricted class. Thus we shall have to make clear what is meant by the addition, multiplication, etc., of real numbers. The definition of these fundamental operations must be such as to be consistent with the definitions of the same processes for rational numbers. We shall need also to see whether the formal laws of arithmetic, such as the commutative and associative laws of addition and the commutative, associative, and distributive laws of multiplication hold equally well for the system of real numbers as for rational numbers. We shall consider these questions in the following articles.

---

[1] The existence of irrational numbers was known to the ancient Greeks and Hindoos. Their discovery was attributed by the Greeks to Pythagorus (about 582–500 B.C.). The complete theory of such numbers was not, however, developed in a satisfactory manner until the last century, when they were considered by Geo. Cantor, Dedekind, and Weierstrass. The first of Cantor's papers on the subject was published in 1871. In 1872 Dedekind published his theory of irrational numbers in his *Stetigkeit und Irrationale Zahlen.*

**4. Cantor's theory of irrational numbers.** Cantor developed the theory of irrational numbers from the properties of regular sequences of rational numbers. Some of the essential features of this theory have already been considered. To complete his theory, it remains to formulate the definitions of such terms as equal, greater than, less than, and of the fundamental processes of arithmetic, based upon regular sequences of rational numbers. As already mentioned, all such definitions and processes must be so formulated as to be consistent with the corresponding definitions already set up for rational numbers.

Given two real numbers $A$ and $B$ defined respectively by the regular sequences $\{a_n\}$ and $\{b_n\}$ of rational numbers. We say that $A$ is **equal** to $B$ and write

$$A = B,$$

if there exists some integer $m$ such that we have for all values of $n > m$

$$| a_{n+p} - b_{n+p} | < \epsilon, \; p = 1, 2, 3, \cdots,$$

where $\epsilon$ is positive and arbitrarily small. The two sequences $\{a_n\}$ and $\{b_n\}$ then define the same number. In order that $A$ and $B$ shall be equal, it is not necessary that $a_n$ and $b_n$ be equal for all values of $n$; for, as we have seen, the same number can be represented by a regular sequence in a great variety of ways.

Of the two numbers $A$ and $B$, $A$ is **greater than** $B$, written $A > B$, if we have for all values of $n > m$

$$a_{n+p} - b_{n+p} > M > 0, \; p = 1, 2, 3, \cdots,$$

where $M$ is a real number independent of $n$. Similarly, $A$ is said to be **less than** $B$, written $A < B$, if we have for $n > m$

$$a_{n+p} - b_{n+p} < M < 0, \; p = 1, 2, 3, \cdots.$$

If $A < B$ and $B < C$, then it follows that $A < C$. As this is true for any three real numbers, the class of real numbers is said to form an ordered system.

The **sum** of the two numbers $A$ and $B$ is defined by the relation

$$\begin{aligned} A + B &= \{a_n\} + \{b_n\} \\ &= a_1 + b_1, a_2 + b_2, a_3 + b_3, \cdot \cdot \cdot, a_n + b_n. \cdot \cdot \cdot \\ &= \{a_n + b_n\}. \end{aligned}$$

By use of this definition the laws governing the addition of rational numbers may be extended to the class of real numbers, including irrational numbers. Thus, the associative law holds, as may be shown as follows: Given the three real numbers $A, B, C$, defined by the regular sequences rational numbers $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, respectively. We have then

$$\begin{aligned} (A + B) + C &= \{a_n + b_n\} + \{c_n\} \\ &= (a_1 + b_1) + c_1, (a_2 + b_2) + c_2, \cdot \cdot \cdot \\ &= a_1 + b_1 + c_1, a_2 + b_2 + c_2, \cdot \cdot \cdot \\ &= a_1 + (b_1 + c_1), a_2 + (b_2 + c_2), \cdot \cdot \cdot \\ &= \{a_n\} + \{b_n + c_n\} \\ &= A + (B + C). \end{aligned}$$

In a similar manner the commutative law of addition may be shown to hold.

We may define the **product** of two numbers $A$ and $B$ represented by regular sequences of rational numbers by the following relation

$$A \cdot B = \{a_n\} \{b_n\} = \{a_nb_n\} = a_1b_1, a_2b_2, a_3b_3, \cdot \cdot \cdot.$$

From this definition the laws of multiplication follow as for rational numbers. We shall show that the distributive law holds and leave the proof for the remaining laws as an exercise for the reader.

Given the three real numbers $A, B, C$, defined respectively by the regular sequences of rational numbers $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, we then have

$$\begin{aligned} (A + B)C &= [\{a_n\} + \{b_n\}]\{c_n\} \\ &= (a_1 + b_1)c_1, (a_2 + b_2)c_2, (a_3 + b_3)c_3, \cdot \cdot \cdot \\ &= a_1c_1 + b_1c_1, a_2c_2 + b_2c_2, a_3c_3 + b_3c_3, \cdot \cdot \cdot \\ &= (a_1c_1, a_2c_2, a_3c_3, \cdot \cdot \cdot) + (b_1c_1, b_2c_2, b_3c_3 \cdot \cdot \cdot) \\ &= AC + BC, \end{aligned}$$

as was to be shown.

From the definitions of addition and multiplication follow at once those of the inverse operations of **subtraction** and **division**. By the multiplication of a number by itself we are led to the integral **powers** of a number, and the inverse operation gives the **extraction of roots.** These definitions of subtraction and division hold for any real numbers, rational or irrational. The definition of a power may be generalized so that it also holds for any rational or irrational exponent. For this purpose consider the numbers $A$ and $\alpha$ defined respectively as the limits of the regular sequences of rational numbers $\{a_n\}$ and $\{\alpha_n\}$. Then by $A^\alpha$ we understand the limit of the sequence

$$\{a_n^{\alpha_n}\} = a_1^{\alpha_1}, a_2^{\alpha_2}, a_3^{\alpha_3}, \cdots,$$

which is seen to be regular. The inverse operation of root extraction may be similarly generalized. It is to be observed, however, that the even roots of negative numbers can not be represented by a regular sequence of rational numbers. These numbers are not real and their consideration involves the introduction of a fundamental unit, $i \equiv \sqrt{-1}$, different from unity. Thus, we see that all of the fundamental operations of arithmetic may be defined for the system of numbers determined by regular sequences of rational numbers; that is, for both rational and irrational numbers, in such a manner that these generalized definitions are consistent with those already employed for rational numbers.

**5. Dedekind's theory of irrational numbers.** In defining real numbers, Dedekind made use of what he called a **partition** of the rational numbers. For example, let $a$ be any rational number. Let the aggregate of all rational numbers be divided into two sets as follows. In the one set $A$ put all those rational numbers which are equal to or less than $a$, and in the other set $B$ all those rational numbers which are greater than $a$. The number $a$ is then said to form a partition of the rational numbers. It serves the purpose equally well if the division is made by putting into the set $A$ the rational numbers less than $a$ and into the set $B$ those equal to or greater than $a$. If the number $a$ is assigned to $A$ then it is the largest number in $A$, and there is no smallest one in $B$; *if* $a$ is assigned to $B$ then it is the smallest number in $B$ and there is no largest one in $A$. In either case, the partition $(A, B)$ is formed by the rational number $a$, and corresponding to every rational number there is such a partition.

We may form a partition $(A, B)$ of the rational numbers such that there is no smallest number in set $B$ and likewise no largest one in set $A$. For example, suppose we put into $A$ the number zero, all negative rational numbers, and all those positive rational numbers whose square is less than 2, and into $B$ those rational numbers whose square is greater than 2. All rational numbers fall into the one set or the other. In this case, there is no largest number in $A$ and likewise no smallest one in $B$. The partition $(A, B)$ may then be said to define the irrational number $\sqrt{2}$.

According to Dedekind the **real number system** consists of that aggregate of objects corresponding to the totality of partitions which may be formed from the system of rational numbers. As any rational number can be so represented, the system of real numbers then includes all rational numbers. Those numbers thus defined that are not rational are called **irrational numbers.** This definition of irrational numbers may be formulated more precisely as follows:

Divide the class of all rational numbers into two sets $A$ and $B$ having the following properties:

1. Every number in $A$ shall be less than any number in $B$, and hence every number in $B$ will be greater than any number in $A$.
2. There shall be no largest number in $A$ and no smallest in $B$.

This partition defines a real number, but it can not define a rational number, since all such numbers are included either in $A$ or in $B$ and none is the last of $A$ or the first of $B$. It must then define an irrational number.

This definition of an irrational number is equivalent to Cantor's definition. To show this, it is necessary to show that every regular sequence of rational numbers may be made to correspond to one and only one partition of the rational numbers, and moreover, that all regular sequences representing the same real number by the Cantor definition correspond to the same partition. Conversely, it is necessary to show that to every partition of the rational numbers there corresponds one and only one regular sequence of rational numbers.

Let $\{a_n\}$ be a regular sequence of rational numbers, thus defining a real number $a$ by the Cantor theory. Let $x$ be a rational number. Consider the regular sequence of rational numbers $\{a_n - x\}$. By Theorem II, Art. 2, the elements of this sequence are ultimately all positive or all negative, provided it does not represent the number zero. Corresponding to the sequence $\{a_n\}$, we form a partition of the rational numbers, if we put into set $A$ all the rational numbers $x$ for which $(a_n - x)$ is ultimately positive, and into set $B$ all those rational numbers $x$ for which $(a_n - x)$ is ultimately negative. If there is a value $x$ for which $(a_n - x)$ is ultimately neither positive nor negative, then assign $x$ to either $A$ or $B$.

Any other regular sequence of rational numbers $\{a_n'\}$ representing the number $a$ must correspond to the same partition as $\{a_n\}$. For, the sequences $\{a_n' - x\}$ and $\{a_n - x\}$ are both convergent and the elements of each must be ultimately either all positive or all negative unless the sequences represent the number zero. Hence, $(a_n - x)$ and $(a_n' - x)$ must have the same sign for all values of $n$ greater than some integer $m$. It follows that if any given rational number $x$ is assigned to the set $A$ in the one case, it will be assigned to the set $A$ in the other. Consequently, we have the same partition of the rational numbers in the two cases.

We shall now consider the converse; namely, we shall show that corresponding to every partition of the rational numbers there exists a regular sequence of rational numbers. Suppose we have given a partition of the rational numbers. Let $A$ and $B$ be the two sets into which the rational numbers are divided. Let $a_1$ and $b_1$ be any two rational numbers in $A$ and $B$, respectively. Then the rational number

$$\frac{a_1 + b_1}{2}$$

lies between $a_1$ and $b_1$. If it is a number of set $A$, then call it $a_2$ and put $b_2 = b_1$. If $\frac{1}{2}(a_1 + b_1)$ is a number of $B$, then call it $b_2$ and put $a_2 = a_1$. In either case

$$\frac{a_2 + b_2}{2}$$

is again a rational number and lies between $a_2$ and $b_2$. If it is a number in $A$, call it $a_3$ and put $b_3 = b_2$; if it is a number in $B$, call it $b_3$ and put $a_3 = a_2$. Take the arithmetical mean of $a_3$ and $b_3$ and proceed as before. Continuing this process indefinitely, we obtain the two sequences of rational numbers

$$\{a_n\} = a_1, a_2, a_3, \cdots, a_n, \cdots,$$
$$\{b_n\} = b_1, b_2, b_3, \cdots, b_n, \cdots.$$

The first of these sequences is monotone increasing and the second is monotone decreasing. Both sequences are bounded, and hence by Theorem III, Art. 2 are convergent. Moreover, they represent the same real numbers; for, we have

$$b_n - a_n = \frac{b_1 - a_1}{2^{n-1}},$$

which is arbitrarily small for sufficiently large values of $n$.

It follows then that corresponding to every regular sequence of rational numbers there exists a partition of the system of rational numbers and conversely. Moreover, the two methods of defining real numbers are equivalent in the sense that they lead to the same system of numbers and any number which is irrational by the one definition is also irrational by the other. It follows that any property or any process which is developed on the basis of the one definition may also be developed on the basis of the other.

**6. The real number system.** The class of real numbers is fully defined by regular sequences of rational numbers or by the partitions of such numbers. The definition of real numbers by either of these methods constitutes an enlargement of the notion of number itself. The introduction of irrational numbers makes possible the use, without restriction, of limits of sequences of rational numbers; for, in the domain of real numbers every regular sequence of rational numbers has a uniquely determined limit which is the real number, rational or irrational, defined by that sequence, a conclusion which is not valid in the domain of rational numbers.

We may speak of sequences of real numbers instead of rational numbers. Moreover, the definition of a convergent sequence can be extended to a sequence of such numbers. Thus, we say that a sequence of real numbers $\{a_n\}$ is convergent, if for an arbitrarily small positive real number $\epsilon$ there exists an integer $m$ such that for all values of $n > m$, we have

$$| a_n - a_{n+p} | < \epsilon, \; p = 1, 2, 3, \cdots .$$

If there exists a real number $A$ such that for all values of $n > m$ we have

$$| A - a_n | < \epsilon,$$

then $A$ is said to be the limit of the sequence $\{a_n\}$ as $n$ becomes infinite, and we write

$$\underset{n \to \infty}{L} a_n = A.$$

Likewise, the arithmetical notions of equal, greater than, less than, as well as the definitions of the fundamental operations of arithmetic may be extended and defined by means of sequences of real numbers instead of rational numbers. All the theorems thus far demonstrated concerning regular sequences of rational numbers can be demonstrated for sequences of real numbers.

Between any two real numbers $a$ and $b$, there exists an irrational number. For, let $\alpha$ be any irrational number, then $\frac{\alpha}{n}$ is also irrational. The real number $a$ is either rational or irrational. Suppose we assume it to be rational. We can so select $n$ that

$$\frac{\alpha}{n} < b - a,$$

and hence we have
$$a < a + \frac{\alpha}{n} < b.$$

But $a + \frac{\alpha}{n}$ is irrational, since it is the sum of a rational and an irrational number. If we assume $a$ to be irrational, we can find a rational number $\beta$ such that

$$\beta < b - a,$$

whence
$$a < a + \beta < b,$$

where $a + \beta$ is an irrational number. As we now see, there always exists between the real numbers $a$, $b$ at least one irrational number.

Likewise, we may show that between any two real numbers $a$, $b$, there exists a rational number. Consider any two rational numbers $\alpha$, $\beta$, where $\alpha < a$, $\beta > b$. Then

$$\gamma = \beta - \alpha, \frac{\gamma}{n}$$

are both rational. By the proper selection of $\kappa$ and $n$, we have

$$a < \alpha + \frac{\kappa\gamma}{n} < b,$$

but $\alpha + \frac{\kappa\gamma}{n}$ is a rational number.

When we considered the totality of sequences of rational numbers, we were led to a new class of numbers, namely irrational numbers. If now we consider all of the regular sequences which may be formed from real numbers, would we not in the same way be compelled to make another extension to our number system? In other words, does every regular sequence of real numbers converge to a real number? To answer this question, we shall make use of the following theorem.

THEOREM. *Let* $\{a_n\}$, $\{b_n\}$, *and* $\{c_n\}$ *be three regular sequences of real numbers such that for all values of* $n$

$$a_n \leqq b_n \leqq c_n,$$

*and if there exists a real number* $A$, *such that*

$$\underset{n\to\infty}{L}\ a_n = \underset{n\to\infty}{L}\ c_n = A;$$

*then we have*
$$\underset{n\to\infty}{L}\ b_n = A.$$

From the given relation

$$a_n \leqq b_n \leqq c_n,$$

we have by subtracting $a_n$ from each member

$$0 \leqq b_n - a_n \leqq c_n - a_n. \tag{1}$$

From the condition that

$$\underset{n \to \infty}{L} a_n = \underset{n \to \infty}{L} c_n = A, \tag{2}$$

we are able to determine an integer $m$ such that for all values of $n > m$ we have

$$c_n - a_n < \epsilon, \tag{3}$$

where $\epsilon$ is arbitrarily small. It then follows from (1) that

$$b_n - a_n < \epsilon, n > m. \tag{4}$$

It follows from the definition of equality that $\{b_n\}$ must then define the same number as $\{a_n\}$, namely $A$, as the theorem states.

In the foregoing theorem the elements $a_n$, $b_n$, $c_n$ of the given sequences were all real numbers. The theorem must then hold for the special cases when the elements of one or all of the given sequences are either rational or irrational numbers. Suppose for example, $a_n$ and $c_n$ are rational numbers, it follows from the definition of real numbers that $A$ is a real number. We may so select $a_n$ and $c_n$ as to define any real number $A$ at pleasure. Since $\{b_n\}$ converges to the same number $A$, it follows that every real number may be represented by means of a regular sequence of real numbers. However, the theorem also holds when $b_n$ is irrational, and hence any real number can also be represented by a regular sequence of irrational numbers. As we have seen, every real number forms a partition of the totality of rational numbers. Let $\{b_n\}$ be any regular sequence of real numbers. If we now take $\{a_n\}$, $\{c_n\}$ to be sequences of rational numbers satisfying the conditions of the foregoing theorem, then the sequence $\{b_n\}$ defines a real number $A$. As $\{b_n\}$ is any regular sequence of real numbers, it follows that every regular sequence of real numbers defines a real number.

The system of real numbers forms a closed system with respect to the fundamental processes of arithmetic, including the extraction of roots of positive numbers and odd roots of negative numbers, and the limiting process; that is, when we operate upon a real number by any one of these processes, we always obtain a real number as a result. Consequently, the system of real numbers thus developed is sufficient for the purposes of a theory of functions of real variables.

Aside from the division of real numbers into rational and irrational, we may make other classifications. For example, if a real number is the root of an algebraic equation

$$a_0x^n + a_1x^{n-1} + \cdot \cdot \cdot + a_{n-1}x + a_n = 0$$

whose coefficients are rational numbers, then the number is called an **algebraic number.** The class of algebraic numbers includes all of the rational numbers and many irrational numbers. For example, the irrational number $\sqrt{2}$ is a root of the quadratic equation

$$x^2 - 2 = 0.$$

The real numbers which are not algebraic are called **transcendental numbers.**[1] The numbers $\pi$ and $e$, the base of the natural logarithmic system, are illustrations of such numbers. Since every rational number can be made the root of an algebraic equation with rational coefficients, it follows that all transcendental numbers are at the same time irrational, although the converse is not true.

To know whether a given number is algebraic or transcendental is often of importance. For example, when it was shown that $\pi$ is a transcendental number, that fact settled a problem which had attracted the attention of mathematicians for centuries; namely, whether it is possible by means of a straight edge and a pair of compasses to construct a square whose area is equal to that of a given circle. For, as may be seen at once, only those numerical values are involved in the constructions made by these instruments which are the roots of an equation either of the first or second degree. Other problems in geometry, such as the trisection of an angle, the construction of a regular polygon of a given number of sides by means of ruler and compasses depend for their solution upon the numerical magnitudes satisfying the foregoing conditions.

**7. Geometric representation of real numbers.** As in analytic geometry suppose we have given a straight line $L$ and on this line a fixed point $O$ called the origin. We shall speak of that portion of the line between any two points $P_1$ and $Q_1$ as the segment $P_1Q_1$. In order to have a basis for geometric interpretation of problems in analysis it is essential that we should have a one-to-one correspondence between the totality of points of a straight line and the totality of real numbers. This relation involves certain assumptions as to the continuity of the points on a straight line and the relation of these points to the system of real numbers.[1] We shall point out what assumptions must be made. The first of these assumptions, known as the **axiom of Archimedes,** may be stated in geometric terms as follows.

---

[1] The existence of transcendental numbers was proved by Liouville in a paper published in *Comptes Rendus* in 1844. The number $e$ was shown to be transcendental by Hermite in 1873, see *Crelle*, Vol. 76, p. 303. Lindemann in 1882 showed that $\pi$ also is transcendental in an article published in the *Mathematische Annalen*, Vol. 20, p. 213. See also article by D. E. Smith in *Monographs on Modern Mathematics* p. 402.

AXIOM I. *Given any two segments* $AB$ *and* $CD$ *of a straight line and suppose* $AB$ *be the shorter; then there always exists an integer* $n$ *such that* $n \cdot AB > CD$.

No matter how short a segment $AB$ may be chosen, the axiom asserts that it is always possible to choose $n$ sufficiently large so that by laying off this segment $n$ times from the point $C$ the final point will be as remote from $C$ as we may care to have it. We make the further assumption that any given segment may be divided into $n$ equal parts, and hence we have upon dividing by $n$ also the relation

$$\frac{CD}{n} < AB.$$

Let the length of the segment $Ox_1$ be taken as the unit of measurement. Then the point $x_1$ can be taken as representing the number 1.

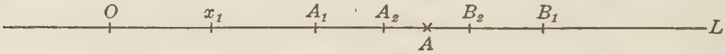


FIG. 1.

By laying this segment off a suitable number of times to the right and to the left of the origin of $O$, we get the points representing the natural numbers. As we have seen, every rational number can be written in the form $\frac{p}{q}$, where $p$ and $q$ are integers prime to each other. By virtue of the foregoing assumption we can subdivide the segment $Ox_1$ into $q$ equal parts. By taking one of these parts as a unit of measure and laying it off $p$ times, beginning at $O$, we obtain the point which represents the rational number $\frac{p}{q}$. The system of rational numbers may therefore be represented uniquely by a set of points on a straight line. We shall speak of these points as the **rational points** of the line. Conversely, to each of these points there corresponds one and only one rational number. There is then a one-to-one correspondence between the rational points of a line and the system of rational numbers.

---

[1] See article by G. Cantor, *Math. Ann.* (1878) Vol. V, p. 28.

When we consider the correspondence between the irrational numbers and the remaining points of the straight line, we need to make a further assumption. The need and the nature of this assumption may be shown as follows. Let $A_1$ and $B_1$ be any two rational points on $L$. Divide the segment $A_1B_1$ into $n$ equal parts. Let $A_2B_2$ be one of these smaller segments. The points $A_2$ and $B_2$ are also rational points, and as before the segment $A_2B_2$ can be divided into $n$ equal parts. Continue this process of division indefinitely and we obtain two sequences of rational points

$$\begin{gathered} A_1, A_2, A_3, \cdots, A_n \cdots, \\ B_1, B_2, B_3, \cdots, B_n \cdots. \end{gathered}$$

Corresponding to these sequences of rational points, we have two regular sequences of rational numbers which we may denote respectively by $\{a_n\}$ and $\{b_n\}$. As $n$ increases indefinitely, the length of the segment $A_nB_n$ becomes less than an arbitrarily chosen value $\epsilon$. We say that the length of the segment approaches zero as $n$ becomes infinite. However, as $n$ becomes infinite the two regular sequences $\{a_n\}$, $\{b_n\}$ define one and the same real number $a$. With Cantor we assume the following axiom.

Axiom II. *On the line $L$ let there be given a sequence of segments $\{A_nB_n\}$, corresponding to the sequences $\{a_n\}$, $\{b_n\}$ of rational numbers, satisfying the condition that each segment lies within the preceding one and that the limit of the length of the segments $A_nB_n$ is zero as $n$ becomes infinite. Then there exists on $L$ one and only one point $A$ which divides all of the rational points $\{A_n\}$ from the rational points $\{B_n\}$.*

This axiom is the geometric analog of the definition of real numbers by means of a partition of the rational numbers or its equivalent, a sequence of rational numbers. It is often referred to as the Cantor-Dedekind axiom or the **axiom of continuity.**

As we have seen, every real number $a$ forms a partition of the aggregate of rational numbers. Corresponding to such a partition of the rational numbers, there exists a sequence of segments $\{A_nB_n\}$ satisfying the conditions of the axiom of continuity. The point $A$ which is the limit of this sequence of intervals may be said to represent the real number $a$. The foregoing axiom being assumed, we may conclude that to every real number there corresponds one and only one point on a straight line.

Conversely, by aid of Axioms I and II it follows that if we have given any point $P$ of a straight line, there corresponds one and only one real number. For, lay off from the origin $O$ the unit segment a sufficient number of times, say $a_1$ times, so that by laying it off once more the terminal point will lie on the opposite side of $P$ from $O$. Let $P_1$ be the point corresponding to $a_1$ and $Q_1$ to $a_1 + 1$. Divide the segment $P_1Q_1$ into $n$ equal parts, which is equivalent to dividing the number one into $n$ equal parts. Taking one of these segments as a measure, lay it off from $P_1$ a sufficient number of times, say $a_2$ times, so that we reach a terminal point $P_2$ on the same side of $P$ as $O$, and laying it off $a_2 + 1$ times we have a terminal point $Q_2$ on the opposite side of $P$ from $O$. The points $P_2$, $Q_2$ then correspond respectively to the rational numbers $a_1 + \frac{a_2}{n}, a_1 + \frac{a_2}{n} + \frac{1}{n}\cdot$ Continuing this process indefinitely, we have the two sequences of rational points

$$P_1, P_2, P_3, \cdots, P_n, \cdots,$$
$$Q_1, Q_2, Q_3, \cdots, Q_n, \cdots$$

and the corresponding sequences of rational numbers; namely,

$$a_1, a_1 + \frac{a_2}{n}, a_1 + \frac{a_2}{n} + \frac{a_3}{n^2}, \cdots,$$
$$a_1 + 1, a_1 + \frac{a_2}{n} + \frac{1}{n}, a_1 + \frac{a_2}{n} + \frac{a_3}{n^2} + \frac{1}{n^2}, \cdots .$$

It follows from the axiom of continuity that either of these sequences of rational points approaches $P$ as a limit, while the corresponding sequences of rational numbers converge and define a definite number $a$.

By aid of the foregoing axioms we have thus set up a one-to-one correspondence between the points of a straight line and the totality of real numbers. Because of this correspondence, all of the properties and definitions stated for the real number system may be restated in geometric terms, and we thus have the possibility of a geometric interpretation of the problems of analysis. In what follows we shall often speak of the points $a$, $b$, etc., meaning those points corresponding to the real numbers $a$, $b$, etc.

The points of the straight line that are not rational are called the **irrational points,** and these points stand in a one-to-one correspondence with the irrational numbers. Whether any particular point is rational or irrational is not an inherent property of the point but rather depends upon the choice of a unit and of the origin. By changing the unit or the origin it may be made a rational or an irrational point at pleasure. The points on the line stand in the same relative order as the corresponding numbers. Thus if the numbers $a, b$ correspond respectively to the points $A$, $B$, then $A$ is to the left of $B$ if $a < b$ and to the right in case $a > b$.

It is possible to show the existence of a one-to-one correspondence between the elements of any finite segment of a straight line and those of a given segment, say (0,1). To do this draw straight lines through the extremities of the given segment (0, 1) and those of the other segment, which is assumed not to lie on the same straight line. The lines through the extremities meet in some point $O$. With this point as a center all of the points of the one segment may be projected uniquely upon the other, thus establishing the desired one-to-one correspondence.

By the properties of harmonic points, we may illustrate how a one-to-one correspondence may be established between the points of a straight line of indefinite length and those of an arbitrarily short segment $AB$. A corresponding analytic proof may be readily supplied by the reader. Upon $AB$ construct a complete quadrilateral as shown in Fig. 2. From the properties of a complete quadrilateral it follows that the points $C$ and $D$ divide the segment $AB$ harmonically. To each position of $C$ there corresponds one and only one point $D$, and conversely to each point $D$ there is one and only one point $C$. As $C$ varies from $A$ to $B$ the point $D$ traverses the entire indefinite line exterior to $AB$, including the point at infinity in a projective sense, and thus the one-to-one correspondence is seen to exist. We may take the segment (0, 1) as the segment $AB$ and hence there exists a one-to-one correspondence between the points of the indefinite line exterior to (0, 1) and those of the segment (0, 1).

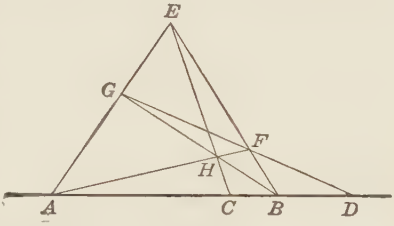


FIG. 2.

We shall speak of the segment of a straight line lying between the points $A$ and $B$ as the **interval** ($A$, $B$). Because of the one-to-one correspondence with real numbers, we shall also speak of the interval ($a$, $b$) meaning the real numbers between $a$ and $b$. If the end-points of the interval are included, we call it a **closed interval,** otherwise it is called an **open interval.** By a **neighborhood** of a point we shall mean an arbitrarily small interval inclosing the point. If the point is not included, we shall speak of it as a **deleted neighborhood.**

Just as we have a sequence of points, so we can have a sequence of intervals. The end-points determine the interval. If we denote by $I_n$ the interval whose end-points are $a_n$, $c_n$, we may state the theorem of Art. 6 in geometric terms as follows.

THEOREM. *Given a sequence* $\{I_n\}$ *of closed intervals such that each lies within the preceding one and suppose*

$$\underset{n\to\infty}{L}\{\textit{Length of } I_n\} = 0.$$

*If* $P_n$ *is any point in the interval* $I_n$, *then every sequence* $\{P_n\}$ *determines one and the same limiting point* $P$.

The proof of this theorem is the precise geometric analogy of that of the theorem in Art. 6 and is left for the reader to supply.

**8. Base of a system of notation.** In the discussion thus far, we have considered only the general case where the unit has been divided into $n$ equal parts. When we come to introduce a particular system of notation, it becomes necessary to give $n$ a definite value called the **base** of the system. Thus, in the decimal system, 10 is the base. The selection of a base determines the number of digits necessary and also certain laws of procedure in expressing numbers in that system. Thus, every removal of a digit one place to the right or left decreases or increases its value $k$ fold if the integer $k$ is selected as a base. We shall now show that any integer greater than one may be selected as the base of a system of notation for real numbers.

Let $A$ be any real number and $k$ any positive integer greater than one. We shall show that we can express the number $A$ in a system of notation having $k$ as a base. In doing so, we shall make use of the geometric representation of real numbers by points on a straight line as a guide in selecting a suitable sequence to define $A$ in the desired manner. Lay off from $O$ the unit segment a sufficient number of times to finally obtain the interval in which the point $A$ lies. If $A$

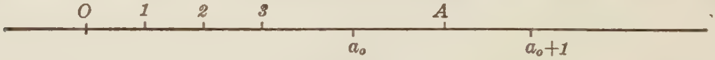


FIG. 3.

is an end-point of this interval, then $A$ is an integer. If it is not, the point $A$ lies within the interval. Suppose it lies between points $a_0$ and $a_0 + 1$. Divide this interval into $k$ equal parts. Suppose that as we proceed from $a_0$ we pass $a_1$ of these divisions before we come to that one in which $A$ lies. Divide this interval into $k$ equal parts and proceed as before, continuing the process indefinitely. The left endpoints of the intervals in the sequence of intervals containing $A$ represent the numbers of the sequence

$$a_0,\ a_0 + \frac{a_1}{k},\ a_0 + \frac{a_1}{k} + \frac{a_2}{k^2},\ a_0 + \frac{a_1}{k} + \frac{a_2}{k^2} + \frac{a_3}{k^3},\ \cdots.$$

This sequence is regular and defines the number $A$ in a system of notation having the base $k$. In such a system of notation there must be $k$ digits.

Suppose for example, $k$ is equal to 3. We must then have three digits 0, 1, 2. Every removal of a digit one place to the left increases its value threefold and likewise every removal one place to the right decreases its value threefold. If now we wish to change some number, say 14, expressed in the decimal system, to the system having the base 3, we write 14 as the sum of several numbers such that the digit in first place is 0, 1, or 2, the digit in the second place some multiple, not to exceed the double, of 3, in the third place a multiple of 9, etc; thus

$$14 = 9 + 3 + 2.$$

In the required system 2 is the same in both systems, 3 is represented by 10, and 9 by 100. Hence the required representation is 112.

## EXERCISES

**1.** Determine which if any of the following sequences are regular:

$$1,\ 1 + \tfrac{1}{2},\ 1 + \tfrac{1}{2} + \tfrac{1}{3},\ 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4},\ \cdots;$$

$$1,\ 1 + \frac{1}{3},\ 1 + \frac{1}{3} + \frac{1}{3^2},\ 1 + \frac{1}{3} + \frac{1}{3^2} + \frac{1}{3^3},\ \cdots;$$

$$2,\ 2 + \frac{1}{2},\ 2 + \frac{1}{2} + \frac{1}{3\cdot 2},\ 2 + \frac{1}{2} + \frac{1}{3\cdot 2} + \frac{1}{4\cdot 2^2},\ \cdots;$$

$$1,\ 1 + \frac{1}{\sqrt{2}},\ 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}},\ 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}},\ \cdots.$$

**2.** Find the limit of the following regular sequence:

$$1,\ 1 + \frac{1}{2},\ 1 + \frac{1}{2} + \frac{1}{2^2},\ 1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3},\ \cdots.$$

**3.** Set up a generalization of the sequence given in Ex. 2 such that it shall represent the number $\dfrac{a}{a-1}$.

**4.** Is the number $a = \log_{10} 5$ a rational or an irrational number?

**5.** Show that the sequence

$$1.9,\ 1.99,\ 1.999,\ 1.9999,\ \cdots$$

defines the number 2. Set up another regular sequence also defining 2. How many such sequences are possible?

**6.** Show that the following sequence converges and defines an irrational number:

$$1, 1 - \tfrac{1}{2}, 1 - \tfrac{1}{2} + \tfrac{1}{3}, 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4}, \cdots.$$

HINT: Compare with the expansion[1] of $\log(1 + x)$ in a Maclaurin series.

**7.** If $\{a_n\}$ defines the number $a$, show that $\{|a_n|\}$ defines the number $|a|$.

**8.** By the method employed in Art. 3 show that $\sqrt{5}$ is an irrational number. Is it also transcendental?

**9.** Define addition of irrational numbers by the use of a partition of the rational numbers and show that the associative law of addition holds.

**10.** Given the segment (0, 1) of a straight line. Divide this segment into three equal parts and drop out the middle part. Divide the remaining segments into three equal parts and drop the middle ones. Continue this process indefinitely. Show that the intervals thus formed may be arranged in sequences so that each sequence defines a single definite point.

**11.** Show that every periodic decimal fraction is a rational number.

**12.** The irrational number $e$ may be defined as the limit as $n$ becomes infinite of the sequence $\left\{\left(1 + \frac{1}{n}\right)\right\}^n$. Prove the convergence of this sequence.

**13.** Show that two regular sequences having an infinite number of elements in common represent the same number.

**14.** Given the number 125 in the decimal notation. Represent the same number in a numeral system having two as the base.

---

[1] See Townsend and Goodenough's *Essentials of Calculus*, p. 307.

## CHAPTER II

## THEORY OF POINT SETS

**9. Infinite aggregates.** In the preceding chapter we have had occasion to consider aggregates of numbers and of points on a straight line. Some of these aggregates, for example the regular sequences of rational numbers, contained an infinite number of elements. One may formulate the properties common to all infinite aggregates and, considered as abstract properties, they may be made the basis of rigorous mathematical discussions irrespective of the character of the individuals composing the aggregate.[1] A scientific statement of the principles which follow from the consideration of these properties is of fundamental importance in mathematics and is known as the theory of aggregates. In what follows, however, we shall direct our attention to those properties which have to do with aggregates of points on a line or in ordinary space of two or three dimensions. Such a restriction of the general theory leads to the theory of **point sets.** If the set of points is confined to a line, we shall speak of it as a **linear set.** In a similar manner, we have a two-dimensional or a three-dimensional set if the points of the set are distributed over the plane, or in ordinary space. We shall also have occasion to consider sets of intervals and of regions. In the following articles we shall consider the general properties of linear sets. Later it will be pointed out how these properties may be extended to sets of two or more dimensions.

We have seen that, assuming certain axioms, we may set up a one-to-one correspondence between the aggregate of all real numbers and the totality of points on a straight line. Consequently, while the discussion will often take a geometric form, nevertheless the substance is always arithmetical, the geometric interpretation being a convenience rather than a necessity. As we shall see, a clear understanding of the properties of point sets is of special value in discussing the nature of the limiting processes of analysis such as continuity, differentiation, integration, etc., and hence to the full understanding of the nature of functions themselves.

---

[1] The properties of infinite aggregates were first studied by Geo. Cantor in 1871. See *Math. Ann.*, Vol. 5.

In passing from the properties of finite aggregates to those of infinite aggregates, we must generalize the fundamental concepts and processes of mathematics and inquire whether the formal laws of arithmetic hold for these generalized notions. It becomes apparent at once that not all of those formal laws can hold for infinite aggregates. For example, in finite aggregates the whole is always greater than any of its parts; that is, we can never put the elements of a part of the aggregate into a one-to-one correspondence with the elements of the whole aggregate. This property does not hold for infinite aggregates. For example, the positive even integers can be put into a one-to-one correspondence with all of the positive integers. Likewise there exists a one-to-one correspondence between the positive integers and those positive integers divisible by three. In fact, the property that the elements of an aggregate may be put into a one-to-one correspondence with those of a part, is characteristic of all infinite aggregates and may be made the definition of such aggregates.

**10. Transfinite numbers.** In elementary mathematics we say that a variable **becomes infinite** if it increases in absolute value without limit; that is, if its absolute value has no finite upper bound. For example, if $n$ becomes infinite we understand that it may take any integral value however large, and we express that fact by writing[1] $n \to \infty$. It should be read "$n$ becomes infinite," or "$n$ increases without limit." The symbol $\infty$ does not denote a definite number, and, in fact, it does not denote a number at all, but is a symbol used to express more conveniently a certain manner of variation. We do not say that $n$ approaches $\infty$, as we say that $\epsilon$ approaches 0, if it decreases indefinitely in numerical value, for the reason that the symbol 0 denotes a definite number and $\infty$ does not.

There is, however, another use of the term infinity which comes into consideration in the higher branches of analysis. With Cantor, we shall assume that the sequence of all positive integers defines a number. Cantor called it the first **transfinite ordinal number,** and denoted it by $\omega$. This number $\omega$ may be said to form a Dedekind partition between the class of all finite numbers and the class of all transfinite numbers. We shall not consider the properties of transfinite numbers, but wish merely to call attention to the possibility of their existence.[1]

---

[1] Often this relation is denoted by writing $n = \infty$.

For finite aggregates, the reader is familiar with the distinction between ordinal numbers and cardinal numbers. The ordinal number of such an aggregate is that number assigned to the last element in the process of counting. This number does not depend upon any particular order or arrangement of the elements. An infinite aggregate may not have a first element nor a last one. In many cases it is, however, possible to place the elements of an infinite aggregate in a definite order of arrangement, so that of any two elements we may say that the one precedes or follows the other according to some law of succession which enables us to say that if $a$ precedes $b$ and $b$ precedes $c$, then $a$ precedes $c$. Two infinite aggregates having the same definite order of arrangement of their elements are said to have the same **ordinal type** if there exists a one-to-one correspondence between their elements such that their order is preserved; that is, if $a_1$ and $a_2$ are elements of the one set which correspond respectively to $a_1'$ and $a_2'$ of the other set and $a_1 < a_2$ (or $a_1 > a_2$), then $a_1' < a_2'$ ($a_1' > a_2'$). Any set of elements which has the same ordinal type as the positive integers is said to have the ordinal number $\omega$. The ordinal number of infinite aggregates differs from that of finite aggregates in that it depends upon a definite arrangement of the elements and may vary if that arrangement is changed. Thus, if we rearrange the positive integers by taking first all of the even integers followed by all of the odd integers, the ordinal number is no longer $\omega$ but $\omega \cdot 2$, since the ordinal number of the even integers and also that of the odd integers is $\omega$.

Again, we need to generalize the notion of a cardinal number in speaking of infinite aggregates. By the **cardinal number** of a finite aggregate is meant the number of objects composing the aggregate irrespective of any order of arrangement; that is, it is the common property of all finite aggregates having a one-to-one correspondence with one another. We may likewise define the cardinal number of an infinite aggregate as the invariant property of the class of all aggregates whose elements can be put into a one-to-one correspondence with those of a given infinite aggregate. The cardinal number of an infinite aggregate does not depend upon any order of arrangement and is of course a transfinite number. We shall have occasion to compare the cardinal numbers of different infinite aggregates. Thus, any aggregate composed of the elements of a regular sequence has the same cardinal number as the set of positive integers, and is usually denoted by the symbol[1] $a$. The set of all real numbers has the cardinal number of the set of points on a straight line, and as this is assumed to be continuous, we say that the set of real numbers has the cardinal number of the continuum, and denote it by the symbol $c$.

---

[1] For a discussion of transfinite numbers, see Hobson's *Theory of Functions of a Real Variable*, Vol. I, Chapters II and IV.

**11. Fundamental definitions.** Among the simplest infinite sets are those whose elements may be put into a one-to-one correspondence with the positive integers. Such a set is said to be **enumerable** or countable.[2] If such a one-to-one correspondence is not possible, then the set is said to be **non-enumerable.** A set composed of the elements of a regular sequence or of the terms of an infinite series are illustrations of enumerable sets. A set composed of a finite number of elements is also enumerable.

A linear set is said to be **bounded** if its points all lie in a finite interval. Similarly, a two or three dimensional set is said to be bounded if its points all lie in a finite region. Thus the linear set

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots, \frac{1}{2^n}, \cdots \tag{1}$$

is bounded, because all of its points lie in the interval (0, 1).

If the points of a set are so numerous in the neighborhood of a point $a$ that in every interval $(a - \epsilon, a + \epsilon)$ however small there are an infinite number of points of the set, then the point $a$ is called a **limiting point** of the set. Thus, in the foregoing set, the point zero is a limiting point, since an infinite number of points of the set lie in any arbitrarily small interval inclosing the origin. We may also express this fact by saying that the given set is **dense** at the origin, or in the general case at the limiting point $a$. If every point of an interval is a limiting point of the given set, then the given set is said to be **everywhere dense** in the interval. Such an interval is characterized by the fact that no subdivision of the interval however small can be found which does not contain points of the given set. For example, the set of rational points is everywhere dense in any given interval; for, every point of the given interval is a limiting point of such a set.

The aggregate of all the limiting points of a given set is called the first **derived set.** In a similar manner the limiting points of the first derived set constitute the second derived set, etc.

---

[1] Also denoted by $\aleph_0$ read "Alef-null."

[2] The term "denumerable" is also frequently used.

We shall distinguish between a limiting point and a **point of condensation.** By a point of condensation, we shall understand a point in every neighborhood of which there are a non-enumerably infinite number of points of the given set. It follows that every point of condensation is a limiting point, but the converse is not true. If the points of condensation themselves have a limiting point, that point is itself a point of condensation, since every such point must have in every neighborhood of it a non-enumerable infinity of points of the given set.

By the **complementary set** of a given set $E$, we mean the set obtained by omitting from the continuum the points of $E$. For example, the irrational points form the complementary set of the rational points. If a given set $A$ contains the set $E$ as a subset, we may also speak of the complementary set of $E$ relative to $A$, by which is meant that set of points obtained by omitting from the set $A$ those points which belong to $E$. Unless otherwise designated, we shall use the term complementary set to mean the complement relative to the continuum. The set complementary to $E$ is denoted by the symbol $C(E)$.

Any point of a set which is not a limiting point is called an **isolated point.** Every such point is characterized by the fact that it may be inclosed in an interval such that no other points of the set lie within this interval. It is possible to have an infinite set such that every point of the set is an isolated point. This is always the case if none of the limiting points of the set are themselves points of the set. Such a set is often spoken of as an **isolated set.** For example, the set

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{n}, \ldots$$

is isolated, since its only limiting point, namely zero, is not a point of the set.

A point $x_0$ is called an **inner point** of a given set $E$ if it may be included in an interval $(x - \rho, x + \rho)$, $\rho \neq 0$, such that all the points of this interval are points of $E$. An inner point is also characterized by the fact that it is a point of $E$ which is not a limiting point of the complementary set $C(E)$. Thus, every point of a segment of a straight line other than the end-points is an inner point of the set composed of all the points of the segment. On the other hand, the set of rational points has no inner points.

A point which can be inclosed in an interval containing no points of the given set is called an **exterior point.** An exterior point is not a point of the given set but is an inner point of the complementary set. A point which is neither an inner point of the set nor an exterior point is called a **boundary point** of the set. The totality of such points constitute the boundary. A boundary point is also characterized by the fact that it is not an inner point of the given set $E$ nor of the complementary set $C(E)$. Such a point is either a point of $E$ which is a limiting point of $C(E)$ or a point of $C(E)$ which is a limiting point of $E$. Thus if $E$ is the set of rational points in the interval (0,1), every point of this interval is a boundary point.

A set which includes all of its boundary points is said to be **closed.** Such a set then contains all of its limiting points, and indeed a closed set might be defined as one which contains all of its limiting points. For, every boundary point is either a point of the set which is a limiting point of the complementary set, or it is a limiting point of the given set. Consequently, if all of the limiting points are included, all of the boundary points are included. The set of all real numbers in a closed interval is an illustration of a closed set.

If every point of a given set is an inner point, then we shall call the set an **open set.**[1] Such a set therefore contains no boundary points. There are then sets which are neither closed nor open according to the foregoing definitions. For example, the set of rational numbers is such a set.

If, in addition to being closed, a set has the property that every point of the set is a limiting point of the set, then it is said to be a **perfect set.** All the points of a segment of a line is a simple illustration of such a set.

By a **continuum** we mean any set which is everywhere dense and perfect. For example, the set of all real numbers or of the points of a straight line form a continuum.

If it is possible to say of any three elements, $a$, $b$, $c$ of an aggregate that if $a$ precedes $b$ and $b$ precedes $c$, then $a$ precedes $c$, the aggregate is said to be **ordered.** The points of a set may be ordered by the process of counting, or they may be arranged in order of magnitude, or by aid of any other suitable property. However, the question of order does not involve the necessity of arrangement by aid of any particular property. It does not involve the possibility of counting the elements and indeed there need be no first element of the set and likewise no last. Thus, the set of all the points on a straight line constitutes an ordered set although there is no first point, and as we shall see later, the points can not be enumerated. If there is a first element of the set and likewise of every subset, we say that the set is **well-ordered.**

---

[1] Some authors define an open set as one which is not closed; for example, see Young's *Theory of Sets of Points*, p. 19.

Two sets are said to be **equivalent** if they have the same cardinal number; that is, if they can be put into a one-to-one correspondence with each other. Thus the two following sets are equivalent:

$$1, 2, 3, 4, \cdot \cdot \cdot , n, \cdot \cdot \cdot ,$$
$$2, 4, 6, 8, \cdot \cdot \cdot , 2n, \cdot \cdot \cdot .$$

It will be at once evident that all infinite enumerable sets are equivalent.

If the points of a given set lie within a finite interval $AB$, the set is said to be bounded. We shall have occasion to speak of the upper and lower limits of a point set and likewise of its upper and lower bounds. Thus, if we have a bounded point set on the $X$-axis, we shall understand by its **upper limit** a definite point $P$ such that there are points of the given set in every neighborhood of $P$ but at most only a finite number of points of the set to the right of any such neighborhood. In other words $P$ is the extreme right-hand limiting point of the given set. The **upper bound** of the set is its upper limit, in case no points lie to the right of $P$, or the extreme right-hand point of the set. Similarly, the **lower limit** of the given set is a definite point $P$ such that points of the set lie in every neighborhood of $P$ but at most only a finite number of points of the set are to the left of any such neighborhood. By the **lower bound** of the set is meant the extreme left-hand point of the set or the lower limit $P$, in case no points of the set lie to the left of $P$. The upper and lower bounds are sometimes spoken of as the least upper bound and the greatest lower bound, respectively. That every bounded set has a least upper bound follows from the fact that the numbers of the linear continuum can be divided into two groups; the first group containing those numbers greater than every number in the set, and the second group containing the numbers of the set and all numbers less than some or all numbers of the set. This partition of the linear continuum defines a real number, which is the least upper bound. Similarly, the existence of the greatest lower bound may be shown. It will be seen that the upper and the lower limits are respectively the upper and the lower bounds of the first derived set; that is, the set composed of the limiting points of the set.

**12. Fundamental operations with point sets.** We shall have occasion to combine in one way or another two or more point sets into a single point set. In this connection it is necessary to define what is to be understood by the terms, sum, product, etc., of two or more point sets. We shall also inquire whether the formal laws of arithmetic hold for those operations.

If we have given two point sets, $E_1$ and $E_2$ such that $E_1$ is contained in $E_2$, we write

$$E_1 < E_2, \text{ or } E_2 > E_1.$$

By the **sum** of the point sets $E_1, E_2, \dots, E_n$, we mean the set $E$ of all those points such that each appears in at least one of the given sets, and we write

$$E = E_1 + E_2 + \cdots + E_n.$$

Although a point may appear in more than one of the sets $E_1, E_2, \dots, E_n$, it is to be taken but once in the sum $E$. It will be seen that $E$ is what is often called the logical sum of the given point sets.[1] The process of addition is commutative and associative as in elementary arithmetic, thus

$$E_1 + E_2 = E_2 + E_1,\ E_1 + (E_2 + E_3) = (E_1 + E_2) + E_3 = (E_1 + E_3) + E_2.$$

Given two sets $E_1$ and $E_2$, the **difference**, $E_1 - E_2$, consists of those points of $E_1$ which are not points of $E_2$. As one sees at once from this definition neither the commutative nor the associative law holds in general for subtraction. Moreover, by this definition subtraction is not the inverse operation of addition for sets having common points.

The **product** of the point sets $E_1, E_2, \dots, E_n$ is the set of points common to all of these sets; that is, it is the logical product of the given sets. If $E$ is the product we write

$$E = E_1 \cdot E_2 \cdot \quad \cdots \quad \cdot E_n.$$

The commutative, associative, and distributive laws hold for multiplication as here defined; that is,

$$E_1 \cdot E_2 = E_2 \cdot E_1,$$
$$(E_1 \cdot E_2)E_3 = E_1(E_2 \cdot E_3),$$
$$E_1(E_2 \pm E_3) = E_1 \cdot E_2 \pm E_1 \cdot E_3.$$

---

[1] See Whitehead and Russell, *Principia Mathematica*, Vol. I, p. 217. The sum is sometimes defined as the combination of two sets having no points in common, while the combination of sets having common points is called the *union* of the given sets. Cantor used the expression least common multiple instead of the union of the given sets. The union of sets having common points is sometimes denoted by $\mathfrak{S}(E_1, E_2, \dots)$ to distinguish it from the sum. Instead of the term product as defined in the text the term *greatest common divisor* is also used.

The difference and the product of two point sets $E_1$, $E_2$ can be expressed as the sum of two point sets if we make use of the complementary sets. For example, the complement of the difference of $E_1$ and $E_2$ is composed of the points of $E_2$ and the complement of $E_1$; that is, we have

$$C(E_1 - E_2) = C(E_1) + E_2. \tag{1}$$

From the definition of a product, it will be seen that the complementary set of the product of $E_1$ and $E_2$ is the set of points not contained in both $E_1$ and $E_2$; that is, it is the sum of the complementary sets of $E_1$ and $E_2$. We may then write

$$C(E_1 \cdot E_2) = C(E_1) + C(E_2). \tag{2}$$

It follows from (1) and (2) that theoretically we need only consider addition and the taking of complements. It will frequently be more convenient, however, to make use of both subtraction and multiplication of sets.

The foregoing discussion has to do with the arithmetical operations on a finite number of point sets. Later we shall see how these operations can be extended and applied to an enumerably infinite number of sets.

**13. Sequence of point sets.** We shall have occasion later to consider a sequence of point sets rather than the individual sets. Suppose we have given the sequence

$$\{E_n\} \equiv E_1, E_2, \cdots, E_n, \cdots,$$

where each $E_n$ is itself a point set. The question presents itself whether such a sequence can have as a limit a definite set of points. In order that such a limit $E$ shall exist, each point $P$ of $E$ must be a point of an infinite number of the sets $E_n$. Moreover, the point $P$ must occur in all of the sets $E_n$ after a certain one. We can best formulate the definition of convergence of a sequence of point sets $\{E_n\}$ by introducing what will be called a complete limit and a restricted limit of the given sequence.

We shall understand by a **complete limit** of $\{E_n\}$ the set of all points such that each occurs in an infinite number of the sets $E_n$. If we denote the complete limit by $E_c$, we may then write

$$E_c = (E_1 + E_2 + E_3 + \cdots)(E_2 + E_3 + \cdots)(E_3 + \cdots)\cdots; \tag{1}$$

that is, $E_c$ must contain the points common to the sum of all the given sets, the sum of all omitting the first, the sum of all omitting the first and second, etc.

The **restricted limit** of $\{E_n\}$ is understood to be the set of all those points which occur in every set $E_n$ after a certain one; that is, there must exist a value of $n$ such that the points in question are contained in all the sets $E_n$, $E_{n+1}$, . . . . If we denote the restricted limit by $E_r$, we then have

$$E_r = E_1 \cdot E_2 \cdot E_3 \cdot \quad \cdot \cdot \cdot + E_2 \cdot E_3 \cdot \quad \cdot \cdot \cdot + \cdot \cdot \cdot ; \tag{2}$$

that is, $E_r$ must contain all of the points common to all of the sets $E_n$, likewise those common to all except $E_1$, those common to all except $E_1$, $E_2$, those common to all except $E_1$, $E_2$, $E_3$, etc.

It follows that all points of $E_r$ are likewise points of $E_c$; that is, $E_r$ is contained in $E_c$. When $E_r$ and $E_c$ are equal, the given sequence $\{E_n\}$ is said to **converge** to unique limiting point set. A sequence $\{E_n\}$ does not necessarily converge, but if it does we may write

$$\underset{n \to \infty}{L} E_n = E_c = E_r = E.$$

From the foregoing definitions it will be seen that the complete limit of the sequence $\{E_n\}$ is the complementary set of the restricted limit of the sequence of complementary sets $C(E_1)$, $C(E_2)$, $\cdot \cdot \cdot \cdot$ .

As an illustration of a convergent sequence of point sets, let us consider the sequence $\{E_n\}$, where

$$E_1 \leqq E_2 \leqq \quad \cdot \cdot \cdot \leqq E_n \leqq \cdot \cdot \cdot , \tag{3}$$

and it is assumed that $E_1$ contains elements; that is, it is not a null set. Every point of any $E_n$ is then contained in all the sets of the sequence which follow. Hence $E_c$ and $E_r$ both exist and $E_c$ is not a null set. Moreover, the two limits are the same. Consequently, the limit exists and we have

$$\underset{n \to \infty}{L} E_n = E_1 + E_2 + \cdot \quad \cdot \cdot + E_n + \cdot \cdot \cdot \cdot .$$

As another illustration of a convergent sequence, consider the sequence $\{E_n\}$, where

$$E_1 \geqq E_2 \geqq \cdot \cdot \cdot \geqq E_n \geqq \cdot \cdot \cdot . \tag{4}$$

Each $E_n$ contains all the sets that follow. Again $E_c$ may be different from the null set, in which case $E_r$ also exists and the two are equal. We, therefore, have the result

$$\underset{n \to \infty}{L} E_n = E_1 \cdot E_2 \cdot \quad \cdot \cdot \cdot \cdot E_n \cdot \quad \cdot \cdot \cdot .$$

Let us now consider any sequence $\{E_n\}$ of point sets. Set up the auxiliary sequence $\{A_n\}$, where

$$A_n = E_1 + E_2 + \cdot \cdot \cdot + E_n.$$

From the first of the above illustrations the sequence $\{A_n\}$ has a limit, and hence we may write

$$E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot = \underset{n \to \infty}{L} (E_1 + E_2 + \cdot \cdot \cdot + E_n). \quad (5)$$

In a similar manner, if we put

$$B_n = (E_1 \cdot E_2 \cdot \ \cdot \cdot \cdot \ \cdot E_n),$$

the sequence $\{B_n\}$ has a limit by the second of the above illustrative examples. We then have as a result

$$(E_1 \cdot E_2 \cdot \ \cdot \cdot \cdot \ \cdot E_n \cdot \ \cdot \cdot \cdot) = \underset{n \to \infty}{L} (E_1 \cdot E_2 \cdot \ \cdot \cdot \cdot \ \cdot E_n). \quad (6)$$

From the definition of a complete limit and of a restricted limit respectively, it is evident that we may write for any set $\{E_n\}$

$$E_c = \underset{n \to \infty}{L} (E_n + E_{n+1} + \cdot \cdot \cdot), \ E_r = \underset{n \to \infty}{L} (E_n \cdot E_{n+1} \cdot \ \cdot \cdot \cdot). \quad (7)$$

**14. Properties of derived sets.** The existence of limiting points of an infinite set is established by the following theorem.

THEOREM I. *An infinite set of points inclosed in a finite interval has at least one limiting point.*

Suppose the given set of points $E$ lies in the finite interval $(a, b)$. Divide this interval into two equal parts by the point $x_1$. Of these two parts at least one, say $(x_1, b)$, must contain an infinite number of points of the given set. Divide the subinterval $(x_1, b)$ into two equal parts by the point $x_2$. At least one of the intervals $(x_1, x_2)$, $(x_2, b)$ must contain an infinite number of points of $E$. Suppose $(x_1, x_2)$ to be the one. Divide this subinterval into two equal parts by the point $x_3$ and proceed as before. Let this process of subdivision be carried on indefinitely. We thus obtain an infinite sequence of intervals $I_n$, such that each lies within the preceding one and the length of $I_n$ approaches zero as $n$ becomes infinite. By the theorem of Art. 7, this sequence defines a definite limiting point $x_0$. In every neighbor-

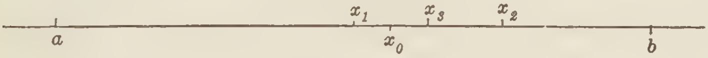


FIG. 4.

hood of $x_0$, however small, there exists an infinite number of points of the given set. The point $x_0$ is, therefore, a limiting point of the set. There may be other limiting points, but the theorem shows that there must be at least one.

In most of the illustrations thus far given, the infinite sets of points have but a single limiting point. That there may be more than one limiting point is illustrated by the following examples.

**Ex. 1.** Consider the limiting points of the set,

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{2}+\frac{1}{3}, \frac{1}{4}, \frac{1}{2}+\frac{1}{4}, \frac{1}{5}, \frac{1}{2}+\frac{1}{5}, \ldots .$$

This set has the limiting point 0 and also the limiting point $\frac{1}{2}$; for, the subset,

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$$

is dense at the origin, and the subset

$$\frac{1}{2}, \frac{1}{2}+\frac{1}{3}, \frac{1}{2}+\frac{1}{4}, \ldots$$

is dense at $\frac{1}{2}$.

**Ex. 2.** Find the limiting points of the point set

$$1, \frac{1}{2}, \left(\frac{1}{2}\right)^2, \frac{1}{2}+\left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \frac{1}{2}+\left(\frac{1}{2}\right)^3, \left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3, \left(\frac{1}{2}\right)^4, \frac{1}{2}+\left(\frac{1}{2}\right)^4,$$
$$\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^4, \left(\frac{1}{2}\right)^3+\left(\frac{1}{2}\right)^4, \left(\frac{1}{2}\right)^5, \ldots .$$

From inspection, it will be seen that this point set has an infinite number of limiting points, namely:

$$0, \frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \ldots .$$

As we have seen, the limiting points themselves constitute a point set called the first derived set. If $E$ is the original set, then the derived set is denoted by $E'$. If $E'$ is composed of an infinite number of points lying within a finite interval, then it also has one or more limiting points and these points constitute the second derived set, which is denoted by $E''$. The point set considered in Ex. 2 has two derived sets, namely

$$E' = 0, \frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \ldots ,$$
$$E'' = 0.$$

If $E''$ had contained an infinite number of points, we should have had a third derived set $E'''$, etc.

If but a finite number of derived sets exist, the given point set is said to be of the **first species.** A point set may, however, have an infinite number of derived sets, and in this case it is said to be a set of the **second species.** The set of all rational numbers in the interval (0, 1) is an illustration of a point set of the second species. In this case, the first derived set contains all the real numbers in the given interval, and each succeeding derived set is identical with the first. We have the following theorem concerning derived sets.

THEOREM II. *Every derived set is closed.*

Let $E$ be the given set, and let us assume that $E'$ is not closed. Then there exists a limiting point $x_0$ of $E'$ which is not a point of $E'$. In every neighborhood of $x_0$ there are then an infinite number of points of $E'$, and each of these points is, therefore, a limiting point of $E$. Hence, in every neighborhood of $x_0$ there exist an infinite number of points of $E$, and $x_0$ must then be a limiting point of $E$ and consequently a point of $E'$, which is contrary to the assumption that $E'$ is not closed. From this contradiction the theorem follows.

**15. Properties of enumerable sets.** The simplest infinite point sets are those whose elements can be placed in a one-to-one correspondence with the positive integers, namely enumerable sets. The division of infinite point sets into enumerable and non-enumerable sets is of great value in discussing many of the important questions arising in analysis and in geometry. Such a point set may be so arranged as to be well-ordered and the elements distinguished from one another by means of subscripts.

Any subset of an enumerable set is also enumerable, and if a finite set is added to an enumerable set the sum is enumerable. Moreover, every enumerable set can be decomposed into an enumerably infinite number of subsets each of which consists of a single point or a finite number of points. Further properties of such sets are given by the following theorems.

THEOREM I. *The sum of an enumerable set of enumerable sets is itself an enumerable set.*

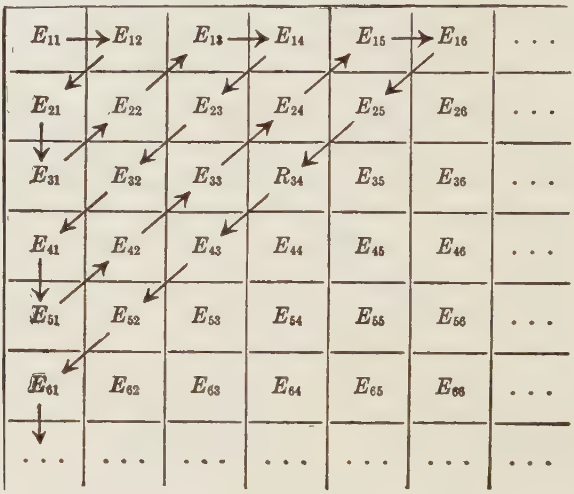
Given the enumerable sets

$$E_1, E_2, \cdot \cdot \cdot, E_n, \cdot \cdot \cdot,$$

each of which consists of an enumerably infinite number of points. We are to show that the sum $E$ is also enumerable. The elements $E_n$ may be written as follows:

$$\begin{array}{l}
E_1 = E_{11},\ E_{12},\ E_{13},\ \cdots,\\
E_2 = E_{21},\ E_{22},\ E_{23},\ \cdots,\\
E_3 = E_{31},\ E_{32},\ E_{33},\ \cdots,\\
\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
E_n = E_{n1},\ E_{n2},\ E_{n3},\ \cdots,\\
\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ .
\end{array}$$

We shall now select the elements of $E$ from this array in such a manner that the sum of the subscripts shall constitute a monotone increasing sequence. This may be done by taking first the element the sum of whose subscripts is 2, followed by those whose sum is 3, then by those whose sum is 4, etc. What is the same thing, we may select them in the order indicated by the arrows in the following arrangement.


$E_{11}$ $E_{12}$ $E_{13}$ $E_{14}$ $E_{15}$ $E_{16}$ . . .
$E_{21}$ $E_{22}$ $E_{23}$ $E_{24}$ $E_{25}$ $E_{26}$ . . .
$E_{31}$ $E_{32}$ $E_{33}$ $R_{34}$ $E_{35}$ $E_{36}$ . . .
$E_{41}$ $E_{42}$ $E_{43}$ $E_{44}$ $E_{45}$ $E_{46}$ . . .
$E_{51}$ $E_{52}$ $E_{53}$ $E_{54}$ $E_{55}$ $E_{56}$ . . .
$E_{61}$ $E_{62}$ $E_{63}$ $E_{64}$ $E_{65}$ $E_{66}$ . . .
. . . . . . . . . . . . . . . . . . . . .

The resulting arrangement of $E$ gives

$$\begin{array}{cccccccc}
E = E_{11}, & E_{12}, & E_{21}, & E_{31}, & E_{22}, & E_{13}, & E_{14}, & \ldots,\\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
1, & 2, & 3, & 4, & 5, & 6, & 7, & \ldots,
\end{array}$$

where the elements stand in a one-to-one correspondence with the positive integers as indicated. The set $E$ is therefore enumerable as the theorem states.

We have the following corollaries.

COR. I. *The sum of a finite number of enumerable sets or the sum of an enumerable set of finite sets is enumerable.*

This corollary follows at once from Theorem I, since every finite set is enumerable, and we then have an enumerable set of enumerable sets.

It follows from this corollary that if an enumerable set is subtracted from a non-enumerable set, the resulting set is necessarily non-enumerable; for, otherwise it would be possible to add two enumerable sets and obtain a non-enumerable set. The difference of two non-enumerable sets may, however, be either enumerable or non-enumerable.

Cor. II. *The set of rational points on a line is enumerable.*

The rational points correspond to the set of rational numbers. Every rational number can be written in the form $\frac{p}{q}$, where $p$ and $q$ are each integers. The set of values which can be assigned to $p$ or to $q$ is therefore enumerable. There is then an enumerable set of such fractions having a given numerator, and since there are an enumerable set of values which can be used as numerators, it follows that the set of all rational numbers constitutes an enumerable set of enumerable sets and hence is, by Theorem I, enumerable.

It is to be observed that the rational numbers when arranged in order of magnitude can not be put into a one-to-one correspondence with the positive integers without disturbing the order of arrangement. When arranged in this manner, there is no next one following a given number; for, as we have seen, there is always an infinite number of rational points between any two given rational points. What the theorem asserts is that it is possible so to arrange the totality of all rational numbers that they are in a one-to-one order-preserving correspondence with the positive integers, and from this fact it follows that they constitute an enumerable set.

It has been pointed out that the set of rational points is everywhere dense, and we now see that in spite of that fact the set is enumerable. It does not follow that the set of all points on a straight line is enumerable. The fact that such a set is not enumerable is shown by the following theorem.[1]

Theorem II. *The linear continuum is a non-enumerable set.*

---

[1] Due to Cantor (1873). This paper is said to contain the first instance where, in dealing with infinite sets, there was need to make use of different cardinal numbers.

Without loss of generality, we can consider the points of the interval between zero and one; for, if the set of points between 0 and 1 is non-enumerable, then the same is true for the entire linear continuum. We shall assume, therefore, that the numbers corresponding to the points of the open interval (0, 1) form an enumerable set and show that this assumption leads to a contradiction. Let these numbers be denoted by

$$P_1, P_2, P_3, \cdots, P_k, \cdots. \tag{1}$$

Each of these numbers can be expressed as a decimal fraction having an infinite number of terms. In case the decimal is terminating, there are two ways in which it can be so represented. Thus the fraction three-tenths can be written either

$$0.3000 \cdots \text{ or } 0.2999 \cdots.$$

To make the representation definite and uniform, we shall write all the terminating decimals with a continuation of ciphers after the last significant figure. The numbers given in (1) may then be written

$$\begin{aligned} P_1 &= 0.\,\alpha_{11}\,\alpha_{12}\,\alpha_{13} \cdots \alpha_{1n} \cdots, \\ P_2 &= 0.\,\alpha_{21}\,\alpha_{22}\,\alpha_{23} \cdots \alpha_{2n} \cdots, \\ P_3 &= 0.\,\alpha_{31}\,\alpha_{32}\,\alpha_{33} \cdots \alpha_{3n} \cdots, \\ &\cdots\cdots\cdots\cdots\cdots \\ P_k &= 0.\,\alpha_{k1}\,\alpha_{k2}\,\alpha_{k3} \cdots \alpha_{kn} \cdots, \\ &\cdots\cdots\cdots\cdots\cdots, \end{aligned} \tag{2}$$

where each symbol $\alpha_{kn}$ in this array represents some one of the digits

$$0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9.$$

We now set up the number

$$P = 0.\,\alpha_1\,\alpha_2\,\alpha_3 \cdots \alpha_n \cdots,$$

where $\alpha_1$ is any of the foregoing digits different from $\alpha_{11}$, $\alpha_2$ is any one different from $\alpha_{22}$, $\alpha_3$ different from $\alpha_{33}$, etc., in general $\alpha_n$ is any digit different from $\alpha_{nn}$, each digit being taken different from 9. We have thus a decimal fraction, and hence $P$ lies in the interval (0, 1). However, the number $P$ is different from any number $P_k$ given in (2), since it differs by at least one digit from each one of them. Consequently, the assumption that the numbers given in (1) correspond to the totality of points in the interval (0, 1) is wrong, since the number $P$ also corresponds to such a point. Hence, the set of all points in (0, 1) is not enumerable.

THEOREM III. *The set of all irrational points on a line is non-enumerable.*

From Theorem II, the set $E_1$ of all points on a line is non-enumerable, and by Theorem I, Cor. II, the set $E_2$ of rational points is enumerable. If we subtract an enumerable set from a non-enumerable set the resulting set is non-enumerable. Hence, $E_1 - E_2$ is non-enumerable as the theorem states.

From this theorem it follows that there are many more irrational points than rational points in a continuum. As the rational and irrational points are everywhere dense, it follows that a set of points which is everywhere dense may or may not be enumerable.

THEOREM IV. *If we have given on a straight line a set of non-overlapping intervals, then the set is enumerable.*

About any point $x_0$ of the given line draw a circle of unit radius. The intervals in question may be very small, but they are always different from zero. Since the lower limit of the lengths of the intervals may be zero, it is possible that an infinite number of these intervals may lie within the circle. However, as the diameter of the circle is equal to two, only a finite number of intervals of length equal to or greater than $\frac{1}{2^n}$ can lie within the circle, whatever finite value be assigned to $n$. Thus there are a finite number having a length between 1 and $\frac{1}{2}$, a finite number of length between $\frac{1}{2}$ and $\frac{1}{4}$, between $\frac{1}{4}$ and $\frac{1}{8}$, etc. The intervals lying within the circle must then form at most an enumerable set of finite sets and hence must be enumerable. To show that the set of intervals exterior to the circle is also enumerable, subject the given line to geometric inversion with respect to the given circle. By this process all of the non-overlapping intervals lying exterior to the circle become non-overlapping intervals within the circle. The set of all such intervals within the circle form at most an enumerable set, as we have seen. It follows that the set of all non-overlapping intervals on the given straight line must be enumerable.

COR. *Every isolated set of points is enumerable.*

Each point of the given set $E$ is an isolated point and hence can be inclosed in an interval containing no other point of the set. If any two of these intervals overlap, one or both may be shortened so that the intervals merely abut. The given set of points is then inclosed in a set of non-overlapping intervals, which is enumerable by the foregoing theorem. As each interval contains but one point of the given set $E$, it follows that $E$ is also enumerable.

THEOREM V. *A set, bounded or not, containing no point of condensation of the set, is enumerable.*

Denote the given set by $E$, and consider the sequence of values

$$\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \cdots, \frac{1}{2^n}, \cdots. \tag{1}$$

The entire line upon which $E$ lies can be covered by an enumerably infinite set of non-overlapping intervals of length less than, or at most equal to, $\frac{1}{2^n}$, $n = 1, 2, 3, \cdots$. As we shall see, these intervals may be used to divide $E$ into an enumerable set of subsets $E_1$, $E_2$, $\cdots$, $E_n$, $\cdots$.

Put into $E_1$ all those points of $E$ which can be inclosed in a set of non-overlapping intervals of length $\frac{1}{2}$ such that each interval contains only an enumerable subset of $E$. Put into $E_2$ all those remaining points of $E$ which can be inclosed in a set of non-overlapping intervals of length $\frac{1}{2^2}$, each interval likewise containing only an enumerable subset of $E$. In general, put into $E_n$ all of those remaining points of $E$ which can be inclosed in intervals of length $\frac{1}{2^n}$, such that the intervals do not overlap with one another nor with those previously used, and each contains only an enumerable subset of $E$. There is thus associated with each number given in (1) at most an enumerable set of non-overlapping intervals.

Since no point of the given set $E$ is a point of condensation, it is possible to assign all of the points of $E$ to some subset $E_n$. Hence, we have all of the points of $E$ included in the enumerably infinite number of enumerable subsets $E_1$, $E_2$, $\cdots$, $E_n$, $\cdots$, and consequently $E$ itself is enumerable by Theorem I.

THEOREM VI. *A non-enumerable set $E$ is composed of an enumerable set $A$ and another set $B$, each point of which is a point of condensation of $B$.*

The points of condensation of a subset are at the same time points of condensation of the given set. If we remove from $E$ its points of condensation, the remaining subset $A$ can have no points of condensation which are points of $A$, and hence by Theorem V it must be enumerable. Since the subset $B$ where, $B = E - A$, differs from $E$

only by an enumerable set $A$, it follows that $E$ must contain the same points of condensation as $B$. Hence, the set $B$ is not only composed of the points of condensation of $E$, but each point of $B$ is also a point of condensation of $B$. We have, therefore,

$$E = A + B,$$

where $A$ and $B$ possess the required properties.

THEOREM VII. *If the first derived set $E'$ is enumerable, the original set $E$ is also enumerable.*

The product $E \cdot E'$ is the set of points common to the sets $E$ and $E'$. The set $E \cdot E'$ is enumerable, since it is a subset of the enumerable set $E'$. The set

$$A = E - (E \cdot E')$$

is composed of isolated points, and hence is enumerable. The set

$$E = A + (E \cdot E')$$

is the sum of two enumerable sets and is therefore enumerable, as the theorem states.

As every derived set is closed, there exists for a set $E$ of the second species certain points which are common to all the derived sets $E^{(n)}$, no matter how large $n$ may be taken. These points form a closed set which we shall denote by $E^{(\omega)}$, where $\omega$ is the first transfinite ordinal number. We may then state the following theorem.

THEOREM VIII. *Every set of the first species, and those of the second species for which the derived set $E^{(\omega)}$ is enumerable, is itself enumerable.*

In case the set is of the first species, there exists a number $r$ such that the $r^{th}$ derived set $E^{(r)}$ contains only a finite number of points. Moreover, each of the sets

$$(E' - E''), (E'' - E'''), \cdot \cdot \cdot, (E^{(r-1)} - E^{(r)}) \qquad (1)$$

is isolated since each is formed by removing from a point set its limiting points, that is its derived set. Each of these subsets is therefore enumerable by the corollary to Theorem IV.

In the case of a set of the second species, we have the enumerably infinite number of sets

$$(E' - E''), (E'' - E'''), \cdot \cdot \cdot, (E^{(r-1)} - E^{(r)}), \cdot \cdot \cdot \qquad (2)$$

plus the enumerable set $E^{(\omega)}$. Here again each of the sets is enumerable. In the first case we had a finite number of enumerable sets, and in the second case we have an enumerably infinite number of enumerable sets and hence in either case the sum $E'$ is enumerable. Consequently, by Theorem VII the given set $E$ must be enumerable.

**16. Open and closed sets.** It follows from the definition of open and of closed sets that the complementary set of a given set is open or closed according as the given set is closed or open. Likewise, the sum of any number of open sets is an open set; for, an inner point of any of the given sets is an inner point of the combined set. It also follows that the product of any number of open sets is an open set, if not a null-set. Similar statements can be made concerning closed sets. Other properties are stated in the following theorems.

THEOREM I. *The points of condensation of a given set form by themselves a closed set.*

Let $E$ be the given set and $E_1$ the set composed of the points of condensation of $E$. Then any limiting point $P$ of $E_1$ must have an infinite number of points of $E_1$ in every neighborhood of it. But each point of $E_1$ is a point of condensation of $E$ and hence has a non-enumerably infinite number of points of $E$ in every neighborhood of it. Consequently, in every neighborhood of $P$ there must exist a non-enumerable number of points of $E$. The point $P$ must then be a point of condensation of $E$ and hence is itself a point of $E_1$. However, $P$ is any limiting point of $E_1$. Consequently, $E_1$ contains all its limiting points and is therefore closed.

THEOREM II. *Let $E$ be any open point set. Then $E$ can be expressed as the sum of a finite or enumerably infinite set of non-overlapping open intervals.*

As the given set $E$ is open, any point $P$ of it can be included in an interval $\delta$ every point of which is a point of $E$. The intervals $\delta$ may be so selected that the end-points of $\delta$ are not points of $E$ but of the complementary set $C(E)$. Because of this fact, the intervals $\delta$ may be assumed to be non-overlapping. From Theorem IV of the previous article, it follows that the intervals $\delta$ form at most an enumerable set. Consequently, the given point set $E$ is the sum of either a finite or enumerably infinite number of non-overlapping open intervals and the theorem is thus established.

We have the following important theorem, known as the **Heine-Borel theorem.**[1]

---

[1] Borel stated the theorem for the case of an enumerable set of intervals. The theorem was extended by Lebesgue to the more general case. See Borel, *Leçons sur la Théorie des Fonctions* (1914) p. 42; also, de la Vallée Poussin. *Intégrales de Lebesgue*, p. 13.

THEOREM III. *Given a set of intervals* $\{I\}$ *such that each point of a given closed interval* $(a, b)$ *is an inner point of at least one of the intervals* $I$. *We may then select from the set of intervals* $\{I\}$ *a finite number having the same property; namely, every point of* $(a, b)$ *is an inner point of some one of them.*

This theorem is equivalent to saying that if a given closed interval can be completely covered by any set of intervals, it can also be covered by a finite set. A set of intervals $\{I\}$ is said to cover the interval $(a, b)$ if every point of $(a, b)$ is a point of some $I$.

To prove the theorem it is sufficient to divide the given interval $(a, b)$ into a finite number of consecutive parts, each of which lies within some interval $I$. For example, let $(a, b)$ be divided into $2^n$ equal parts. We shall show that there exists a finite value of $n$ such that each subdivision lies within some interval $I$. Let us assume that the contrary is the case and show that this assumption leads to a contradition of the hypothesis. Under this assumption there must exist for each $n$ a subdivision $\delta_n$ which does not lie within any $I$ no matter how large $n$ be taken, and hence we have a sequence of such intervals $\{\delta_n\}$ each lying with the preceding one. As $n$ increases, the length of $\delta_n$ grows smaller and gives rise to a sequence of values having the limit zero. Let $P_n$ be any point in $\delta_n$. We then have for increasing $n$ an infinite set of points, which must have a definite limiting point. This limiting point $P$ must lie in each interval $\delta_n$. As the points of $\delta_n$ do not lie within any $I$ interval, it follows that $P$ can not be an inner point of any $I$. This is impossible, since by hypothesis $P$ is a point of $(a, b)$ and hence must be an inner point of some $I$.

As a consequence of the foregoing theorem, we have the following corollary.

COR. *If we can cover the whole of an interval* $(a, b)$ *by the aid of an enumerably infinite set of intervals* $\{I_n\}$, *then the sum of the lengths of these intervals is at least equal to* $(a, b)$; *that is,*

$$\Sigma \text{ (Length of } I_n) \geqq b - a.$$

The Heine-Borel theorem may be stated in a more general form for a closed set of points as follows.

THEOREM IV. *Given a set of intervals* $\{I\}$ *such that each point of a given bounded, closed set of points* $E$ *is an inner point of at least one of the* $I$ *intervals, then it is possible to select a finite number of the* $I$ *intervals such that each point of* $E$ *is likewise an inner point of some one of them.*

In some discussions this form of the theorem will be found more convenient than that given in Theorem III. The proof is essentially like that given for Theorem III.

It can be shown that the foregoing theorem holds if the set of intervals $\{I\}$ is replaced by an aggregate of open sets.[1] The Heine-Borel theorem can also be extended to two or more dimensional sets.

**17. Perfect sets.** As already stated, a perfect set contains its own limiting points and every point of the set is a limiting point of points of the set; that is, the set is closed and dense in itself. The set of all real numbers is an illustration of a perfect set. Among the properties of perfect sets are those given in the following theorems.

THEOREM I. *A perfect set is identical with its derived set.*

This theorem follows at once from the definition of a perfect set. Since the given set $E$ is dense in itself, it can contain no points which are not at the same time points of the derived set $E'$. Moreover, as $E$ is closed, all of the limiting points, that is all of the points of $E'$, are also points of $E$. Hence, $E$ and $E'$ are identical as the theorem states.

It is to be noted that a perfect set can have no isolated points.

COR. *A perfect set is identical with any one of its derived sets.*

This result follows directly from Theorem I; for, if the derived set $E'$ is identical with $E$, then $E'$ is also a perfect set and must be identical with its first derived set $E''$, etc.

THEOREM II. *If a set $E$ is dense in itself, then its first derived set $E'$ is perfect.*

Denote by $A$ the set of limiting points of $E$ which are not included in $E$. Then the set $B$ given by

$$B = E + A$$

is a perfect set, since it is not only dense in itself but contains all of its limiting points. By Theorem I it follows that

$$B = B' = E'.$$

Hence, $E'$ is also perfect, since it is identical with $B$.

THEOREM III. *The points of condensation of a given non-enumerable set form by themselves a perfect set.*

By Theorem VI, Art. 15, it follows that the given set $E$ can be decomposed into two subsets, $A$ which is enumerable and $B$ of which each point is a point of condensation of $B$. The points of condensation of $E$ form a closed set by Theorem I, Art. 16, and consist of the points of $B$ plus their limiting points. However, every point of $B$ is a limiting point of $B$, and hence the points of condensation of $E$ constitute the derived set $B'$. As $B$ is dense in itself, $B'$ is also, and moreover $B'$ is a closed set. Consequently, $B'$ is a perfect set, and the theorem follows.

---

[1] See de la Vallée Poussin, *Intégrales de Lebesgue,* p. 112.

Since a perfect set contains its own limiting points and each point of the set is such a point, one might infer that the set must be everywhere dense. This, however, is not true as shown by the following illustration.

Consider the interval (0, 1). Divide this interval into three equal parts and drop out the middle third, which we shall designate by $\delta_1$.

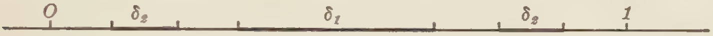


FIG. 5.

Divide each of the two remaining portions likewise into three equal parts and drop out the middle third. We shall denote each of the intervals thus dropped by $\delta_2$. Divide each of the four remaining intervals into three equal parts and as before drop the middle third in each case. Continue this process indefinitely. Let the set $E$ be the end-points of the set of intervals $\{\delta_n\}$ dropped together with the limiting points of these end-points.

It follows from the manner in which points of $E$ are selected that each point is a limiting point of the set. The given set is then dense in itself. Moreover, it is also closed, since by definition $E$ contains all of its limiting points. Hence, $E$ is a perfect set.

However, this set of points is not in any subinterval everywhere dense; for, at each point of $E$ which is an end-point the points of $E$ are dense on only one side. Moreover, in the neighborhood of a limiting point of such end-points there is a portion free from points of $E$ because of the fact that it is a limiting point of end-points.

THEOREM IV. *Every linear perfect set which is in no interval everywhere dense consists of the end-points of an enumerable set of non-overlapping intervals together with their limiting points, and conversely every set of points so defined is a perfect set.*

As the given set $E$ is closed, there are two points of the set which are the extremities of the segment $I$, within which $E$ must lie. The complementary set $\{c\} \equiv C\ (E)$ is an open set and by Theorem II, Art. 16 can be expressed as the sum of an enumerable set of non-overlapping open intervals $\{\delta_k\}$. We shall speak of the open intervals $\delta_k$ as the free intervals, as they contain no points of $E$. Since $E$ is in no subinterval of $I$ everywhere dense, it follows that in every subinterval of $I$ there are points of $\{c\}$. There are then intervals $\delta_k$ which lie wholly or partly within each subinterval of $I$, and consequently the set of intervals $\{\delta_k\}$ is everywhere dense. Since every point $c$ of the complementary set $C\ (E)$ is a point of one of the free intervals $\delta_k$ and these intervals are everywhere dense, it follows that the points of $E$ consist of the end-points of the intervals $\delta_k$ and the limiting points of such end-points, as the theorem requires.

The converse of the given theorem may be readily established, and is left for the reader to supply.

THEOREM[1] V. *A linear perfect set has the cardinal number of the continuum.*

Let us consider the most unfavorable case; namely, where the given perfect point set $E$ is in no subinterval everywhere dense. If we establish the theorem for this case, it holds for all other cases.

We shall first show that to each point of $E$ there may be made to correspond one and only one point of the continuum. By Theorem IV, the set $E$ consists of the end-points of an enumerable set of non-overlapping intervals $\{\delta_k\}$ together with their limiting points. As in the proof of that theorem, these intervals may be so chosen as to be everywhere dense in the interval $I$ containing $E$. In the interval $I$ let us consider also an enumerable, everywhere dense set of points $X$ so situated that any two points have the same relative position as the corresponding intervals.

Let $l$ be the largest $\delta_k$ interval in $I$. We have then

$$I = \delta_0' + l + \delta_1',$$

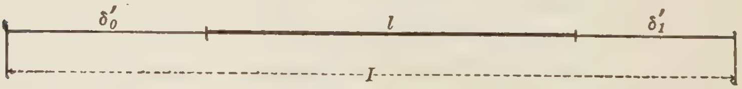


FIG. 6.

where $\delta_0'$ is the interval to the left and $\delta_1'$ the one to the right of $l$. Denote by $l_0$ the largest $\delta_k$ interval in $\delta_0'$ and by $l_1$ the largest $\delta_k$ interval in $\delta_1'$. We may then write

---

[1] See Cantor, *Math. Ann.* Vol. 23, p. 480.

$$\delta_0' = \delta_{00}' + l_0 + \delta_{01}',$$
$$\delta_1' = \delta_{10}' + l_1 + \delta_{11}',$$

where $\delta_{00}'$, $\delta_{10}'$ are to the left of $l_0$, $l_1$, respectively, and $\delta_{01}'$, $\delta_{11}'$ are to the right. Continue indefinitely this process of subdivision of the $\delta'$ intervals. Consider any sequence of intervals of the type

$$\{\delta_N'\} \equiv \delta_p', \delta_{pq}', \delta_{pqr}', \cdots, \delta_N', \cdots \tag{1}$$

selected from the $\delta'$ intervals. The digits $p, q, r, \cdots$ are either 0 or 1. Let $x_N$ be any point of $X$ in the interval $\delta_N'$. There is then associated with the sequence of intervals (1) a sequence of points, namely

$$\{x_N\} \equiv x_p, x_{pq}, x_{pqr}, \cdots, x_N, \cdots. \tag{2}$$

Each interval $\delta_N'$ lies in the preceding one and has one end-point in common with it. Moreover, the lengths of these intervals approach zero as a limit, and consequently (1) determines uniquely a point which is either one of the end-points of a $\delta_k$ interval or a limiting point of those end-points, according as the digits in the subscript

$$N = pqr \cdots$$

are or are not ultimately all 0's or all 1's. But, as already pointed out, the end-points of $\delta_k$ together with their limiting points constitute the given set $E$, and consequently every point of $E$ is the limit of some sequence $\{\delta_N'\}$. On the other hand, the corresponding set of points $\{x_N\}$ has as a limiting point a definite point $P$ of the continuum. Hence, to every point of $E$ corresponds one and only one point of the continuum.

In order to complete the demonstration, we must show, conversely, that to each point of the linear continuum there corresponds one and only one point of the given perfect set $E$. To do so, consider again an enumerable set of points $X \equiv \{x_k\}$ which is everywhere dense in the given interval $I$. Consequently, the first derived set $X'$ is identical with the continuum. For convenience, we shall put

$$X' = \{x'\} = X + X_l,$$

where $X_l$ denotes those limiting points of $X$ not contained in $X$. The points of every convergent sequence $\{x_n\}$ selected from $\{x_k\}$ can be written in the form $\{x_N\}$ given in (2); for, all one needs to do is to express the numbers $n$ in a dyadic number system. The limiting point $x'$ is then a point of $X$ or of $X_l$, according as the digits in

$$N \equiv pqr \cdots$$

are or are not ultimately all 0's or all 1's. However, the sequence $\{x_N\}$ may be made to correspond to the sequence of intervals $\{\delta'_N\}$ having the same subscripts.

As we have seen, the sequence $\{x_N\}$ determines a point $x'$ which is either a point of $X$ or a point of $X_l$. In the first case, it corresponds to an end-point of one of the intervals $\delta_k$, and in the second to a limiting point of such end-points. Hence, to each point $x'$ of the continuum there must correspond one and only one point of $E$. With this conclusion, the demonstration is completed.

The foregoing theorem has some important consequences, among which is the following corollary.

COR. *A perfect set can not be enumerable and conversely.*

By Theorem II, Art. 15, the continuum is non-enumerable. Since we have now shown that the points of any perfect set have a one-to-one correspondence with the points of the continuum, it follows at once that a perfect set is also non-enumerable, in spite of the fact that these points may not be in any subinterval everywhere dense.

The converse also follows from the foregoing theorem; for, if the given set is enumerable, it cannot be put into a one-to-one correspondence with the points of the continuum and can not be perfect.

THEOREM VI. *A closed set is either enumerable, or is composed of an enumerable set plus a perfect set.*

If the given set $E$ is non-enumerable, then by Theorem VI, Art. 15, it follows that it is composed of an enumerable set plus a set $B$, every point of which is a point of condensation. However, $E$ is closed and hence contains all of its own limiting points and consequently all of its own points of condensation. But by Theorem III, the points of condensation of a non-enumerable set form by themselves a perfect set. Hence, $B$ is a perfect set and $E$ is either enumerable or composed of an enumerable set and a perfect set, as the theorem states.

**18. Point sets of the first and second category.** Consider an enumerable set of point sets

$$A_1, A_2, A_3, \cdots, A_n, \cdots, \tag{1}$$

each of which is in no part of a given interval or region everywhere dense. The set of points $E$ formed by the sum of the sets (1), namely

$$E = A_1 + A_2 + \cdots + A_n + \cdots, \tag{2}$$

is said to be a set of the **first category.** Each point of such a set is then a point of at least one of the sets in (1), but no point is taken more than once in making up the set $E$. Every set which is not of the first category is said to be of the **second category.**

The set of all rational points in any interval $(a, b)$ form a point set of the first category. For, we may take as $A_n$ all of those fractions having $n$ as the denominator. Since all rational numbers may be written in the form $\frac{p}{q}$, it follows that the set $E$ defined by (2) is composed of an enumerable set of point sets, each of which is not everywhere dense in any part of $(a, b)$.

The following theorems, while discussed with special reference to linear sets, hold also for two dimensional sets.

THEOREM I. *Every enumerable set is of the first category.*

It follows from the fact that the given set is enumerable that it can be represented as the sum of an enumerably infinite number of sets, each of which consists of a single point. Since this is true, the given set satisfies the definition of a set of the first category.

The converse of this theorem is not true; that is, there are point sets of the first category which are non-enumerable. The following illustration shows that sets of the first category exist which are not only non-enumerable but also everywhere dense in a given interval and moreover have the cardinal number of the continuum.

In Art. 17 it was shown that we may have in the interval $(0, 1)$ a perfect set which is in no subinterval everywhere dense. The perfect set was defined as the end-points of a set of intervals $\{\delta_n\}$, together with their limiting points, obtained by dropping the middle third of the successive intervals. Let $E_1$ denote the point set thus obtained. Consider now each interval $\delta_n$ and treat it in precisely the same manner that the original interval was treated. In this way we obtain a new set of points which added to $E_1$ constitute a perfect set which we shall denote by $E_2$. Treating in the same way those remaining intervals whose inner points are not points of $E_2$ and continuing the process indefinitely, we have a sequence of perfect point sets

$$E_1, E_2, \cdots, E_n, \cdots,$$

each of which includes the previous one, and none of which is everywhere dense in the given interval $(0, 1)$. The sum of these point sets, namely

$$E = E_1 + E_2 + \cdots + E_n + \cdots,$$

is therefore a set of the first category. The resulting set $E$ is, however, everywhere dense in $(0, 1)$; for, in $E_1$ the maximum distance between points of the set is $\frac{1}{3}$, in $E_2$ it is $\frac{1}{3^2}$, in $E_n$ it is $\frac{1}{3^n}$, etc. Hence, as $n$ takes increasing values this maximum distance decreases and becomes arbitrarily small. Since $E_n$ is a perfect set for each value of $n$, it follows that $E$ is also a perfect set. By Theorem V of Art. 17, it follows that $E$ has the cardinal number of the continuum, and hence can not be enumerable.

Theorem II. *The sum of a finite or enumerably infinite number of sets of the first category is a set of the first category.*

Let the given sets whose sum is to be considered be

$$E_1,\ E_2,\ E_3,\ \cdots,\ E_k,\ \cdots.$$

Each of these sets is assumed to be the sum of an enumerably infinite set of other sets, each of which is in no part of the given interval $(a, b)$ everywhere dense. The sets $E_k$ may be represented respectively as follows:

$$\begin{array}{l}
E_1 \equiv A_{11} + A_{12} + A_{13} + \cdots + A_{1n} + \cdots, \\
E_2 \equiv A_{21} + A_{22} + A_{23} + \cdots + A_{2n} + \cdots, \\
E_3 \equiv A_{31} + A_{32} + A_{33} + \cdots + E_{3n} + \cdots, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
E_k \equiv A_{k1} + A_{k2} + A_{k3} + \cdots + A_{kn} + \cdots, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots.
\end{array}$$

We now put

$$B_1 = A_{11},\ B_2 = A_{12},\ B_3 = A_{21},\ B_4 = A_{31},\ \cdots,$$

$B_k$ being one of the elements $A_{mn}$, $k = m + n$, chosen in the same order as in the proof of Theorem I, Art. 15. As each $B_k$ is in no part of the interval $(a, b)$ everywhere dense, the sum of all of the $B$'s, that is

$$E = B_1 + B_2 + \cdots + B_k + \cdots,$$

satisfies the definition of a set of the first category.

We have seen that a point set of the first category may be everywhere dense. The complementary set is always everywhere dense as shown by the following theorem.

Theorem III. *If $E$ is a set of the first category in the interval $(a, b)$, then $C(E)$ is everywhere dense in $(a, b)$.*

The given set $E$ is formed from an enumerably infinite number of sets

$$A_1,\ A_2,\ \cdots,\ A_n,\ \cdots,$$

each of which is in no part of $(a, b)$ everywhere dense. If $(a_0\, b_0)$ is any subinterval of $(a, b)$, it then follows that there is some portion of $(a_0, b_0)$, say $(a_1, b_1)$, which is free from points of $A_1$. Likewise $A_2$ can not be everywhere dense in $(a_1, b_1)$, so there must be some portion of this interval, say $(a_2, b_2)$, which is free from the points of $A_2$. Continuing in this manner we get a sequence of intervals

$$(a_0, b_0),\ (a_1, b_1),\ (a_2, b_2),\ \cdots,\ (a_n, b_n),\ \cdots,$$

each of which is contained in the previous intervals. This sequence defines either a definite point $x_0$ according to the theorem of Art. 7 or an interval whose length is different from zero. There is then at least one point in $(a_0, b_0)$ which is not a point of $E$ and hence belongs to $C(E)$. As $(a_0, b_0)$ is any subinterval of the given interval, it follows that points of $C(E)$ are to be found in every subinterval however small of $(a, b)$; that is, $C(E)$ is everywhere dense in $(a, b)$, as the theorem states.

As an immediate consequence of this theorem, we have the following corollary.

COR. *The continuum is a set of the second category.*

The complementary set to the continuum contains no points and hence cannot be a set which is everywhere dense. Therefore, by the foregoing theorem the continuum cannot be of the first category; for, if it were of the first category, its complementary set would be everywhere dense. Since it can not be of the first category, it must be of the second category.

THEOREM IV. *The complementary set of one of the first category is of the second category.*

From Theorem III, the complementary set is in every portion of the given interval everywhere dense. Denote the given set by

$$E = A_1 + A_2 + \cdots + A_n + \cdots.$$

Assume the complementary set to be also of the first category, and denote it by

$$C(E) = P_1 + P_2 + \cdots + P_n + \cdots,$$

where $P_n$ $(n = 1, 2, 3, \cdots)$ is in no portion of the given interval everywhere dense. The sum

$$E + C(E) = (A_1 + P_1) + (A_2 + P_2) + \cdots + (A_n + P_n) + \cdots$$

is likewise a set such that each term $(A_n + P_n)$ is in no subinterval everywhere dense. This sum must then be a set of the first category, but this is impossible since it is identical with the continuum.

If the given set is of the second category, it does not follow that the complementary set is of the first category; for, there are sets of the second category whose complementary sets are also of the second category.[1]

Theorem V. *Every set of the second category, which is complementary to one of the first category, has the cardinal number of the continuum.*[2]

As we have seen, a set $E$ of the first category may or may not be everywhere dense in the interval containing it. Without loss of generality, we may assume this interval to be (0, 1). The complementary set $C(E)$ is everywhere dense in the given interval (0, 1) by Theorem III. We shall undertake to show that there is a one-to-one correspondence between the points $C(E)$ and those of the continuum (0, 1).

Let $t$ be any point in the open interval (0, 1). Then the number can be written in the dyadic system of numbers as follows

$$t = 0.x_1x_2x_3 \cdots x_n \cdots,$$

where the digits $x_n$ are either 0 or 1. The given set $E$ can be expressed as the limit of the sequence $\{E_n\}$, where none of the sets $E_n$ are everywhere dense in any portion of (0, 1) although $E$ itself may be everywhere dense in that interval. Let $\delta$ be any subinterval of (0, 1) containing no points of $E_1$; that is, it contains only points of $C(E_1)$. Within $\delta$ may be found two subintervals $\delta_0$, $\delta_1$ such that each contains only points of $C(E_2)$. Within each of these intervals can be found two subintervals, denoted respectively by $\delta_{00}$, $\delta_{01}$ and $\delta_{10}$, $\delta_{11}$, such that each contains only points of $C(E_3)$. Continuing this process indefinitely we obtain various sequences of intervals of the type

$$\delta_{x_1}, \delta_{x_1x_2}, \cdots, \delta_N, \cdots,$$

where $N = x_1x_2x_3 \cdots x_n$ and each digit $x_n$ is either zero or one. Each of these intervals lies within the preceding, and they may be so chosen that the length of $\delta_N$ approaches zero as $n$ becomes infinite. Hence, $\{\delta_N\}$ determines in the limit a single point of $C(E)$. Each point $t$ corresponds to a definite sequence of intervals $\{\delta_N\}$, namely that one where $\underset{n \to \infty}{L} N$ is the sequence of digits defining $t$. Conversely, every point of $C(E)$ can be made the limit of a sequence of intervals of the type $\{\delta_N\}$, and to each such sequence there corresponds a definite point of (0, 1). Moreover, the correspondence is unique, if we exclude those sequences in which the digits $x_n$ are from some point on all 1's; for, otherwise two sequences $\{\delta_N\}$ might correspond to the same value of $t$, since

---

[1] See Lebesgue, *Journal de Math.*, Vol. 70 (1905), p. 186.

[2] A set of the second category which is complementary to one of the first category is sometimes called a **residual set**. See Denjoy, *Journal de Math.* (1915), Vol. 80, p. 123.

$$0.\,x_1x_2 \cdots x_n0111 \cdots,$$
$$0.\,x_1x_2 \cdots x_n1000 \cdots$$

represent the same value of $t$. Hence, to each point $t$ there corresponds one and only one point of $C(E)$, and conversely to each point of $C(E)$ there is one and only one point $t$ of the interval (0, 1), and the theorem follows.

THEOREM VI. *Given a finite or enumerably infinite number of sets of the first category. The product of their complementary sets is the complement of a set of the first category.*

Denote the given sets by

$$\begin{aligned} E_1 &= A_{11} + A_{12} + \cdots + A_{1n} + \cdots, \\ E_2 &= A_{21} + A_{22} + \cdots + A_{2n} + \cdots, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ E_m &= A_{m1} + A_{m2} + \cdots + A_{mn} + \cdots, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{3}$$

where each $A_{mn}$ is in no subinterval of the given interval $(a, b)$ everywhere dense. The product of the complements is composed of those points which do not enter into any $E_m$; that is, we have

$$[C(E_1) \cdot C(E_2) \cdots \cdot C(E_n) \cdot \cdots] = C(E_1 + E_2 + \cdots + E_n + \cdots).$$

Consider the sequence of sets $\{B_n\}$, formed by taking the elements of (3) in the order of their occurrence in successive diagonals; that is,

$$B_1 = A_{11},\ B_2 = A_{12},\ B_3 = A_{21},\ B_4 = A_{31},\ B_5 = A_{32}, \cdots.$$

Each of these sets is in no subinterval of $(a, b)$ everywhere dense, and hence their sum, namely

$$E = B_1 + B_2 + \cdots + B_n + \cdots$$

is of the first category. The complementary set $C(E)$ is by Theorem IV of the second category. It is, moreover, the complementary set of one of the first category, a statement that cannot always be made of a set of the second category.

**19. Non-linear sets.** Thus far we have considered only linear sets. Many of the properties developed may be at once extended to the case where the points are distributed over the plane or through space. We shall now consider some of these extensions particularly to two-dimensional sets. The definition of such fundamental terms as limiting point, point of condensation, everywhere dense, etc., may be extended at once to non-linear sets. For interval we substitute region and speak of a two-dimensional set as **bounded** if it is bounded in each dimension; that is, if all the points lie within a definite rectangle. Just as in the case of linear sets, every bounded two-dimensional infinite set has at least one limiting point. Moreover, every sequence of rectangles, such that each lies within the preceding and the limit of the areas is zero, defines in the limit a segment of a straight line or a definite point according as one or both of the dimensions approach zero. By the **distance** between two points whose coördinates are $(x_1, y_1)$, $(x_2, y_2)$ is understood the positive value of $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. By the distance between two sets is meant the greatest lower bound of the distances between the points of one set and those of the other. For many purposes it is convenient to make use of what Jordan calls[1] the **écart,** namely

$$|x_2 - x_1| + |y_2 - y_1|,$$

instead of the distance.

One of the important relations between linear and non-linear sets is given in the following theorem.

Theorem I. *There exists a one-to-one correspondence between the totality of points of a plane and those of the linear continuum.*

This theorem is equivalent to saying that the two-dimensional continuum has the same cardinal number as the linear continuum. As already pointed out, the points on an indefinite straight line may be placed in a one-to-one correspondence with the points of segment $(0, 1)$. This can be done on both axes of coördinates, and consequently we can establish the theorem by showing a one-to-one correspondence between the points of the interval $0 < x \leqq 1$ and the unit square, where $0 < x \leqq 1$, $0 < y \leqq 1$, as indicated in Fig. 7.

Let $P(x, y)$ be any point in this square. The coördinates $x, y$ are real numbers and can be represented in a number system having the base 2. We may then write

---

[1] See *Cours d'Analyse,* Vol. I, p. 23; also Fréchet, *Rend. Palermo,* Vol. 22, p. 53.

$$x = 0.x_1x_2 \cdot \cdot \cdot x_k \cdot \cdot \cdot ,$$
$$y = 0.y_1y_2 \cdot \cdot \cdot y_k \cdot \cdot \cdot ,$$

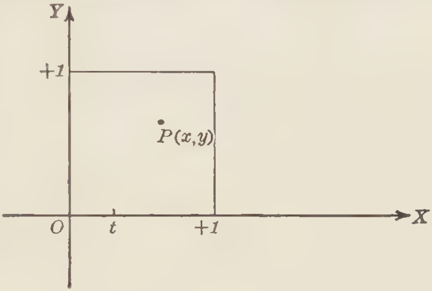


FIG. 7.

where $x_k$, $y_k$ are either 0 or 1. In order to avoid ambiguity, we shall assume that the digits defining $x$, $y$ contain the digit 1 an infinite number of times. With this understanding, the coördinates of $P(x, y)$ can be expressed in one and only one way as a dyadic fraction. Divide the digits composing these two decimal fractions into groups so that each group ends with the digit 1. Thus, if $x$ and $y$ have the particular values

$$x = 0.1\,0\,0\,1\,0\,1\,1\,0\,0\,0\,1 \cdot \cdot \cdot ,$$
$$y = 0.0\,1\,0\,1\,1\,0\,0\,1\,0\,0\,1 \cdot \cdot \cdot ,$$

the arrangement by groups is

$$x = 0.(1)\ (001)\ (01)\ (1)\ (0001) \cdot \cdot \cdot ,$$
$$y = 0.(01)\ (01)\ (1)\ (001)\ (001) \cdot \cdot \cdot .$$

From this group arrangement, we form a new decimal number $t$ by taking a group of digits first from one of the numbers $x$, $y$ and then from the other. Thus, in the particular case just given we have

$$t = 0.101001010111001 \cdot \cdot \cdot .$$

The number $t$ thus defined is represented by a definite point of the segment (0, 1), and thus to any point $P$ of the square there corresponds one and only one point $t$ of the required segment.

Conversely, given any point $t$ of the segment (0, 1) represents uniquely a dyadic number, with the foregoing restriction as to the digit 1 contained in it. We can show by reversing the foregoing process that there corresponds one and only one point $P$ of the given unit square. Since the required correspondence exists, the theorem follows.

The Heine-Borel theorem may be stated for two dimensional space as follows.

THEOREM II. *Let $R$ be a closed rectangular region and let $E$ be a set of rectangular regions such that any point of $R$ is an inner point of some one of the regions in $E$. Then we may select from $E$ a finite number of regions such that any point of $R$ is also an inner point of one of them.*

The proof of this theorem is similar to that given for the corresponding theorem for linear sets.

**20. Content of point sets.** Among the most important properties of point sets are those which concern the notions of content and of measure. These notions are of fundamental value in many of the modern discussions in analysis, particularly in questions concerning differentiation and integration. In fact Hankel[1] first discussed the question of content in connection with the problem of integration. Cantor[2] later developed the general theory. The theory thus developed, however, had certain defects when applied to non-closed sets. Peano[3] and Jordan[4] have undertaken to remedy these defects by introducing what they have called inner and outer content. The notion of content as originally defined by Hankel and Cantor is now of interest mostly for its historical importance. It has been largely replaced by the notion of measure to be defined later. In the discussions to follow, we shall, however, use the term content to mean what is sometimes called the Jordan *measure* or *extent* of a point set.[5] Stated for linear sets this definition may be formulated as follows:

Let us suppose that the given set $E$ lies in a definite interval $(a, b)$. Divide $(a, b)$ into a finite number of subintervals whose lengths may be denoted by $\delta_1, \delta_2, \cdots, \delta_k, \cdots, \delta_n$, where $n$ has any positive integral value. Denote by $\delta$ the norm of the $\delta_k$'s; that is, the maximum of the values $\delta_1, \delta_2, \cdots, \delta_n$. Some of these intervals may contain only inner points of $E$, while others may contain no such points, and still others may contain points of the set and points exterior to the set. All intervals containing boundary points belong to this last class. We are interested in the first and last of these classes of intervals. Denote by $\Sigma_1$ the sum of the lengths of those intervals $\delta_k$ which contain only inner points of $E$, and by $\Sigma_2$ the sum of the lengths of those intervals which contain both points of $E$ and of the complementary set.

Suppose the number of intervals into which $(a, b)$ is divided be increased indefinitely in such a manner that the norm $\delta$ approaches zero. It may be shown[6] that $\Sigma_1, \Sigma_2$ approach definite limits which are independent of the manner in which the given interval is subdivided into the $\delta_k$ intervals. We shall denote these limits by $\Sigma_1'$ and $\Sigma_2'$, respectively. The limit $\Sigma_1'$ is called the **inner content** of $E$. On the other hand, $\Sigma_1' + \Sigma_2'$ is called the **outer content** of $E$.

---

[1] *Math. Ann.* (1882), Vol 20, p. 87.

[2] *Math. Ann.* (1884), Vol. 23, p. 453.

[3] *Applicazioni geometriche del calcolo infinitesimale* (1887), p. 153.

[4] *Journal de Math.* (4) Vol. 4, p. 77 (1892); *Cours d'Analyse*, Vol. 1, pp. 28, 29.

[5] Jordan used the term *eténdue*, and German writers commonly use the term *Inhalt* as the equivalent.

[6] See Jordan, *Cours d'analyse*, Vol. I, pp. 28–31.

If the inner content and outer content are the same, that is if $\Sigma_2'$ is zero, then $E$ is said to be measurable in the Jordan sense and we shall speak of this common value as the **content** or **Jordan measure** of the given set.

It follows from this definition that every set which does not have a portion which is identical with the continuum must have the inner content zero. Such a set is measurable in the foregoing sense only if the outer content is also zero. Hankel and Cantor used the term content as identical with outer content as just defined.

The following is a simple illustration of a measurable set. Given

$$E = 1, \frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \quad . \quad . \quad . \quad , \frac{1}{2^n}, \quad . \quad . \quad . \tag{1}$$

There is but one limiting point, namely 0, which is not a point of the set, and hence is a boundary point. The inner content is zero. The outer content is also zero; for, if we inclose the point zero in an arbitrarily small interval, there remains only a finite number of points of $E$ exterior to it, each of which can be included in an arbitrarily small interval. Hence, all the points of $E$ can be included in a finite number of intervals the sum of whose lengths is arbitrarily small. As the inner content and the outer content are both zero, it follows that the given point set is measurable having the content zero.

The rational numbers in the interval (0, 1) form a non-measurable set in the Jordan sense; that is, this set does not have a definite content. The inner content is zero, since there are no subintervals however small that contain only inner points of the set. On the other hand, the outer content is one, since the set is everywhere dense.

The definition of content can be extended to space of any number of dimensions. For example, let $E$ be a set of points in two-dimensional space. Suppose the plane to be covered by a grating formed by drawing lines parallel to the axes of coördinates as shown in Fig. 8. Denote by $a$ the length of one side of these squares. Some squares may contain only inner points of $E$. Other squares may contain boundary points of $E$. Denote by $\Sigma_1$ the sum of those squares which contain only inner points of $E$, and by $\Sigma_2$ the sum of those squares which contain points of $E$ and points of $C(E)$. Let the number of squares increase indefinitely by making the length $a$ of one side approach the limit zero. As in the case of linear sets, $\Sigma_1$ and $\Sigma_2$ approach definite limits as $a$ approaches zero. Denote these limits by $\Sigma_1'$, $\Sigma_2'$, respectively. The outer content is $\Sigma_1' + \Sigma_2'$, while $\Sigma_1'$ is the inner content. If the two limits are equal, that is if $\Sigma_2'$ is 0, then the given set is said to be measurable in the Jordan sense.

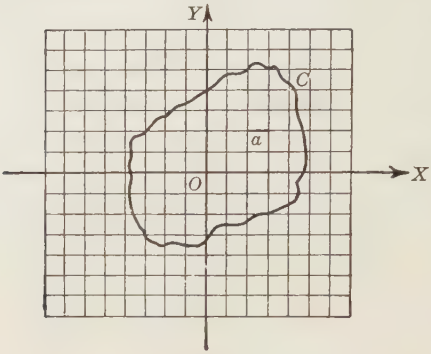


FIG. 8.

If a point set has the content zero, it is sometimes called a **discrete set.** The set given in (1) is an illustration. It will be seen that a set can not be discrete unless its outer content is zero. Moreover, the inner content must also be zero and this can never be the case if in any portion of the given interval the points of the set form a continuum.

Among the theorems involving content, we have the following:

THEOREM I. *The content of a set $E$, which is measurable in the Jordan sense, is equal to the content of its derived set $E'$.*

First of all we shall show that the outer contents of $E$ and $E'$ are the same. Since by hypothesis $E'$ has a definite content, we can inclose the points of $E'$ in a finite number of intervals the sum of whose lengths is arbitrarily small, and the number of points of $E$ exterior to those intervals is finite; for, otherwise there must be at least one limiting point of $E$ and hence a point of $E'$ which has not been included. Inclose each of these $n$ isolated points in an interval of arbitrarily small length $\epsilon$. The outer content of $E$ is then equal to the outer content of $E'$ plus $n \cdot \epsilon$. But $n$ is finite and $\epsilon$ is arbitrarily small, and consequently the sum of the lengths of the intervals inclosing the points of $E$ which are not points of $E'$ is arbitrarily small; that is, the outer contents of the two sets differ by less than any arbitrarily small amount by the proper choice of $\epsilon$ and hence are equal.

We shall show that the inner contents of the two sets are also equal. By hypothesis $E$ is measurable in the Jordan sense, and hence its outer and inner contents are the same. The inner content of $E'$

can not be less than that of $E$, for $E'$ must contain all subsets of $E$ which consist of only inner points and therefore form by themselves a continuum. In other words, if there is any subinterval such that all of its points are points of $E$, then that subinterval appears in the sum of the intervals inclosing points of $E'$. Moreover, the inner content of $E'$ can not exceed the outer content of $E'$. But, as we have seen, the outer contents of $E$ and $E'$ are the same, and this common value must be equal to the inner content of $E$. Consequently, the inner content of $E'$ is less than or at most equal to that of $E$. It follows from the fact that the inner content of $E'$ can not exceed that of $E$ and can not be less than it, that the inner contents of $E'$ and $E$ must be equal.

Since the outer and the inner contents of $E$ are respectively equal to those of $E'$, it follows that the content of $E$ and $E'$ are the same, as the theorem states.

The first part of the foregoing demonstration shows that every open set has the same outer content as that closed set of which it is a part; for, the closed set is none other than the first derived set plus at most a set whose outer content is zero. In the discussions to follow, particularly in the consideration of the Riemann integrals, we shall have occasion to consider the properties of discrete sets, that is sets whose content is zero. It follows also that if the first derived set is discrete, then the given set is discrete.

THEOREM II. *Every bounded point set of the first species is discrete.*

The given set $E$ can have but a finite number of derived sets. Let $E^{(k)}$ be the last derived set. Then $E^{(k)}$ contains only a finite number of points; for, otherwise it would have at least one limiting point and $E^{(k+1)}$ would be different from zero, which is contrary to the assumption that $E^{(k)}$ is the last derived set. Inclose each of the points of $E^{(k)}$ in an arbitrarily small interval of length $\epsilon_1$. If there are $n_1$ such points, then the sum of the lengths of the intervals inclosing these points is $n_1\epsilon_1$. Each of these intervals incloses an infinite number of points of the derived set of next lower order and hence of the original set. Exterior to these $n_1$ intervals there are only a finite number of points belonging to $E^{(k-1)}$. Denote by $n_2$ the number of such points. As before inclose these points in an arbitrarily small interval of length $\epsilon_2$. Their sum is therefore $n_2\epsilon_2$. Exterior to the $n_1 + n_2$ intervals thus far considered there are but a finite number of points, say $n_3$, of the derived set $E^{(k-2)}$. Inclose these in intervals of length $\epsilon_3$ and proceed as before. Continuing this process, after $k + 1$ operations we have all of the points of the given set inclosed in intervals of arbitrarily small length. We may now put

$$\epsilon = n_1\epsilon_1 + n_2\epsilon_2 + \cdot \cdot \cdot + n_i\epsilon_i + \cdot \cdot \cdot + n_{k+1}\epsilon_{k+1},$$

where each term is arbitrarily small upon the proper choice of $\epsilon_i$. Consequently, $\epsilon$ is also arbitrarily small, and we have all the points of $E$ inclosed in a finite number of intervals the sum of whose lengths is arbitrarily small. It follows that the outer content of $E$ is zero.

The inner content of $E$ is also zero; for, the inner content is never negative and can not exceed the outer content. As both the inner and the outer content of $E$ is zero, it follows that its content is zero and the set is discrete.

There are point sets of the second species which are not discrete. The continuum is a simple illustration. On the other hand, there are sets of the second species which are discrete. For example, it is possible to have a perfect set of content zero. The following is an illustration.

Let the required set $E$ be defined as in Art. 17 by means of a set of intervals $\{\delta_n\}$ which is enumerable and everywhere dense in the interval (0, 1). The length of $\delta_n$ is $(\frac{1}{3})^n$ and we exclude the middle third of each interval by the previous subdivision, as indicated in Fig. 9. We regard the $\delta$ intervals as open intervals; that is, we regard the end-points as points of the remaining intervals. The set $E$ is defined as the set of end-points of the $\delta$ intervals increased by their limiting points. The outer content of $E$ is the limit of the sum of the remaining free intervals. This outer content is zero, provided we can show that the limit of the sum of the $\delta$ intervals is equal to the length of the original interval, namely 1. This sum is the limit of the series.

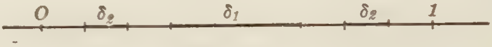


FIG. 9.

$$\frac{1}{3} + 2\left(\frac{1}{3}\right)^2 + 4\left(\frac{1}{3}\right)^3 + \cdot \cdot \cdot + 2^{n-1}\left(\frac{1}{3}\right)^n + \cdot \cdot \cdot$$
$$= \frac{1}{3} + \frac{1}{2}\left\{\left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^3 + \cdot \cdot \cdot + \left(\frac{2}{3}\right)^n + \cdot \cdot \cdot\right\}$$
$$= \frac{1}{3} + \frac{1}{2}\left(\frac{2}{3}\right)^2\left\{1 + \left(\frac{2}{3}\right) + \cdot \cdot \cdot + \left(\frac{2}{3}\right)^{n-2} + \cdot \cdot \cdot\right\}$$
$$= \frac{1}{3} + \frac{2}{9}\,\frac{1}{1 - \frac{2}{3}} = \frac{1}{3} + \frac{2}{3} = 1.$$

Since the limit of the sum of the $\delta$ intervals is 1, the limit of the sum of the remaining intervals, that is the outer content of $E$, is 0.

The inner content of $E$ is also 0, as it cannot exceed the outer content. The content of $E$ is then zero, since both the outer and the inner content is zero. But, as we have already seen, the set $E$ is a perfect set and consequently is of the second species.

Among the additional properties of discrete sets we have the following.

THEOREM III. *Any bounded closed set is discrete, provided it is enumerable.*

Since the given closed set is enumerable, it can be inclosed in an enumerably infinite set of intervals such that each point is an inner point and the length of each interval is arbitrarily small. For example, let each point $E_n$ of the given set $\{E_n\}$ be inclosed in an interval whose length is $\frac{\epsilon}{2^n}$, where $\epsilon$ is arbitrarily small. The sum of these intervals, namely

$$\frac{\epsilon}{2}+\frac{\epsilon}{2^2}+\cdots+\frac{\epsilon}{2^n}+\cdots=\epsilon\left[\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^n}+\cdots\right]=\epsilon,$$

is then also arbitrarily small. By the Heine-Borel theorem, it follows that a finite number of these intervals can be selected such that all of the points of the given set are inner points of these intervals. As each interval is of length $\frac{\epsilon}{2^n}$, the sum of a finite number of them is arbitrarily small, and hence the outer content of the given set is zero. The inner content is also zero, since it can not exceed the outer content, and the theorem follows.

It is of interest to note that in case the given set is non-enumerable, its content may be zero or anything between zero and any value which is less than the length of the interval containing the set.[1]

THEOREM IV. *A discrete set can not be everywhere dense in any subinterval.*

By Theorem I the content must be the same as that of the derived set. If the given set is everywhere dense, then the first derived set is a portion of the continuum itself and of course has a content different from zero.

The converse of this theorem is not true; for, there exist point sets which are in no subinterval everywhere dense and still are not discrete.

---

[1] See Harnack, *Math. Ann.*, Vol. 19, p. 239. The example given by Osgood, *Amer. Math. Jour.*, Vol. 19, p. 167, is also an illustration.

THEOREM V. *Let $E_1, E_2, \cdots$ be a sequence of bounded, discrete sets having the property that each is contained in the one that follows, and let $E$ be the limiting set of the sequence. Then any closed subset of $E$ is discrete.*[1]

The set $E$ is characterized by the properties that every point of $E$ is a point of some set $E_n$ of the sequence, and conversely every point of any set $E_n$ in the sequence is also a point of $E$. Since each $E_n$ is discrete, its points can be enclosed in a finite number of intervals the sum of whose lengths is less than $\frac{\epsilon}{2^n}$, where $\epsilon$ is an arbitrarily small number.

It follows that all the points of the sequence $\{E_n\}$, and consequently those of any subset of $E$, can be made the inner points of a set of intervals, the sum of whose lengths is less than

$$\Sigma \frac{\epsilon}{2^n} = \epsilon \left\{ \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} + \cdots \right\} = \epsilon.$$

As each $E_n$ is bounded, the set $E$, and hence any subset of $E$, is bounded. It follows from the Heine-Borel theorem that any closed subset of $E$ can be inclosed in a finite number of the above intervals and hence has the outer content zero. As the inner content must also be zero, it follows that the subset is discrete, as the theorem states.

**21. Measure of point sets.** The notion of content, as defined in the previous article, has certain disadvantages when applied to point sets in general. In fact, there exists a large class of non-measurable sets, if we base the classification on the definition of content. As a basis for a more general theory, Borel[2] and Lebesgue[3] have introduced what they call the measure of a point set. It differs essentially from the definition of content in that the points of the given set are inclosed in a finite or an enumerably infinite number of non-overlapping open intervals, the sum of whose lengths is finite, instead of being inclosed in a finite number of intervals. Of the two definitions of measure, that of Lebesgue will be made use of in the discussions to follow unless otherwise stated.

The given set $E$ is assumed to be contained in a finite interval $(a, b)$. The intervals used to inclose the points of $E$ may be selected in an infinite number of ways. By each choice the sum of the lengths of the intervals is a definite number. The aggregate of all these numbers has a greatest lower bound, which is called the **exterior measure** of $E$, and is denoted by $m_e(E)$. It is always possible, therefore, to select an enumerable set of intervals $\{\delta_n\}$ inclosing the points of $E$ so that

---

[1] See W. H. Young, *Quarterly Journal*, Vol. 35, (1904), p. 189.

[2] See *Leçons sur la théorie des fonctions* (1898), p. 46.

[3] See *Ann. di Math.* (1902), series III, Vol. 7, p. 231; also *Leçons sur l'intégration*, p. 102.

$$\Sigma\, \delta_n < m_e(E) + \epsilon, \tag{1}$$

where $\epsilon$ is arbitrarily small.

If we denote the length of the interval $(a, b)$ by $l$, we define the **interior measure** of $E$ as

$$l - m_e[C(E)],$$

and denote it by $m_i(E)$. If the interior measure and the exterior measure are the same, we say that the set is measurable, and we call this common value the **measure** of the given set. It is clear that if a given bounded set is measurable, then the complementary set $C(E)$ is also measurable.

We shall understand the measure of an interval to mean its length, and consequently if any set of points $G$ is composed of a set of intervals $\{\ \delta\ \}$, we may write

$$m(G) = \Sigma\, m\delta.$$

In the interval $(a, b)$, let us now consider an enumerably infinite number of point sets $G_1, G_2, \cdots, G_n, \cdots$, each of which is composed of a finite or enumerably infinite number of sets of non-overlapping intervals. Any two intervals may have a common end-point, which belongs to one or the other of the intervals but not to both. The aggregate of intervals composing the sum of the point sets $G_n$ is also enumerable, since it consists of an enumerably infinite number of enumerable sets of intervals. The intervals of this aggregate will overlap only if two or more of the sets $G_n$ have points in common. In case no two of the sets $G_n$ have points in common, we have

$$m(G_1 + G_2 + \cdots + G_n + \cdots) = m(G_1) + m(G_2) + \cdots + m(G_n) + \cdots. \tag{2}$$

If, however, two or more of the sets $G_n$ do have points in common, we have

$$m(G_1 + G_2 + \cdots + G_n + \cdots) \leqq m(G_1) + m(G_2) + \cdots + m(G_n) + \cdots. \tag{3}$$

If two sets, say $G_1, G_2$, have points in common, we have also

$$m(G_1 + G_2) = m(G_1) + m(G_2) - m(G_1 \cdot G_2). \tag{4}$$

This relation can, of course, be extended to the sum of any finite number of the sets $G_n$.

As we shall now see, the foregoing properties of the sets $G_n$ can be employed in studying the measure of open sets and also closed sets of points. By Theorem II, Art. 16, an open set of points can be expressed as the sum of a finite or enumerably infinite number of open intervals. The results given in (2), (3), and (4) then apply to all open sets of points. Consequently, if we have a finite or enumerably infinite number of open point sets $A_n$, no two of which have points in common, we have from (2)

$$m(A_1 + A_2 + \cdots + A_n + \cdots) = m(A_1) + m(A_2) + \cdots + m(A_n) + \cdots. \qquad (5)$$

If two or more of the sets $A_n$ do have points in common, then from (3) we have

$$m(A_1 + A_2 + \cdots + A_n + \cdots) \leqq m(A_1) + m(A_2) + \cdots + m(A_n) + \cdots. \qquad (6)$$

Any two open sets $A_1$, $A_2$ having points in common satisfy the relation given in (4), and hence we have

$$m(A_1 + A_2) = m(A_1) + m(A_2) - m(A_1 \cdot A_2). \qquad (7)$$

In case $A_1$, $A_2$ have no points in common, then $m(A_1 \cdot A_2)$ is zero.

From these properties of open sets, we are able to deduce the corresponding properties for closed sets; for, the measure of a closed set $E$ is the length of the interval $(a, b)$ containing it diminished by the measure of the complementary set $C(E)$, that is

$$m(E) = (b - a) - m[C(E)], \qquad (8)$$

where $C(E)$ is necessarily an open set since it is the complement of the closed set $E$.

If we have given two closed sets $E_1$ and $E_2$ whose complementary sets are respectively the open sets $A_1$ and $A_2$, we have

$$C(E_1 \cdot E_2) = A_1 + A_2,$$
$$C(E_1 + E_2) = A_1 \cdot A_2.$$

From (8) we obtain

$$\begin{aligned} m[C(E_1 \cdot E_2)] &= m(A_1 + A_2) = (b - a) - m(E_1 \cdot E_2), \\ m[C(E_1)] &= m(A_1) = (b - a) - m(E_1), \\ m[C(E_2)] &= m(A_2) = b - a - m(E_2), \\ m[C(E_1 + E_2)] &= m(A_1 \cdot A_2) = (b - a) - m(E_1 + E_2). \end{aligned}$$

Substituting these values in (7), we get

$$m(E_1 + E_2) = m(E_1) + m(E_2) - m(E_1 \cdot E_2). \tag{9}$$

If the given sets $E_1$, $E_2$ have no points in common, then $m(E_1 \cdot E_2)$ vanishes and we have

$$m(E_1 + E_2) = m(E_1) + m(E_2). \tag{10}$$

The exterior measure is always equal to or greater than the interior measure. For, let the given set $E$ be inclosed in the intervals $\{\delta\}$, while those of the complementary set are inclosed in a set of intervals $\{\gamma\}$. It follows from the corollary to the Heine-Borel theorem that

$$\Sigma\{\delta\} + \Sigma\{\gamma\} \geqq b - a.$$

Expressed in terms of measure, we have

$$m_e(E) + m_e[C(E)] \geqq m(a, b),$$

or

$$m_e(E) \geqq m(a, b) - m_e[C(E)] = m_i(E).$$

If the exterior measure is zero, then it follows that the interior measure is also zero and the set is measurable, having the measure zero. The exterior measure has certain other properties which we shall have occasion to use. For example, if we have given two sets $E_1$, $E_2$, where $E_1 > E_2$, we have at once the relation

$$m_e(E_1) \geqq m_e(E_2). \tag{11}$$

If we have given a finite or enumerably infinite number of sets of points $E_1, E_2, \cdots, E_n, \cdots$ contained in an interval $(a, b)$, then

$$m_e(E_1 + E_2 + \cdots + E_n + \cdots) \leqq m_e(E_1) + m_e(E_2) + \cdots + m_e(E_n) + \cdots. \tag{12}$$

For, each set $E_n$ can be inclosed in a finite or enumerably infinite set of open intervals

$$\delta_{n1}, \delta_{n2}, \delta_{n3}, \cdots, \delta_{nk}, \cdots.$$

Consequently, all of the points of the sequence $\{E_n\}$ can be inclosed in an enumerably infinite number of enumerably infinite sets of open intervals which, we know, is an enumerable set of open intervals. We shall denote this set of intervals by $\{\delta_n\}$. Since $m\Sigma\delta_n$ is the length of the sum of these intervals, we have

$$m_e(E_1 + E_2 + \cdots + E_n + \cdots) \leqq m(\delta_1 + \delta_2 + \cdots + \delta_n + \cdots).$$

As the sum of the lengths of the intervals $\delta_n$ is at least equal to the length of their sum, this inequality can be written

$$m_e(E_1 + E_2 + \cdots + E_n + \cdots) \leqq m(\delta_1) + m(\delta_2) + \cdots + m(\delta_n) + \cdots.$$

The greatest lower bound of the sum of the lengths, that is the measures of the intervals $\delta_n$, is the sum of the exterior measures of the sets $E_n$. Hence, the intervals $\delta_n$ may be so chosen that

$$m(\delta_1) + m(\delta_2) + \cdots + m(\delta_n) + \cdots = \epsilon + m_e(E_1) + m_e(E_2) + \cdots + m_e(E_n) + \cdots,$$

however small the positive number $\epsilon$ may be. Hence, from the above inequality, we have

$$m_e(E_1 + E_2 + \cdots + E_n + \cdots) \leqq \epsilon + m_e(E_1) + m_e(E_2) + \cdots + m_e(E_n) + \cdots.$$

Since this relation holds for every arbitrarily small positive value of $\epsilon$, we have in the limit the required relation given in (12).

As already mentioned, the definition of measure enables us to include among the measurable sets a much larger number than does the definition of content. For example, we saw that the set of rational numbers in the interval (0, 1) was not measurable in the Jordan sense. It is, however, measurable in the Borel-Lebesgue sense; for, it is an enumerable set and all such sets are measurable, as we shall now show.

Suppose the given set $E$ lies in the interval (0, 1). When arranged in an ordered one-to-one correspondence with the positive integers, we can inclose the first point of the set in an open interval of arbitrarily small length $\epsilon$. The second point can then be inclosed in an open interval of length $\frac{\epsilon}{2}$, the third in one of length $\frac{\epsilon}{2^2}$, and in general the $n^{th}$ in an open interval of length $\frac{\epsilon}{2^{n-1}}$, etc. Thus, all of the points of $E$ can be inclosed in a set of non-overlapping open intervals whose sum does not exceed

$$\begin{aligned}&\epsilon + \frac{\epsilon}{2} + \frac{\epsilon}{2^2} + \cdots + \frac{\epsilon}{2^{n-1}} + \cdots \\ &= \epsilon\left\{1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} + \cdots\right\} = 2\epsilon.\end{aligned}$$

As $\epsilon$ is arbitrarily small, it follows that $2\epsilon$ is also, and hence the given set can be inclosed in an enumerably infinite set of open intervals the sum of whose lengths has the lower limit zero; that is, the exterior measure of $E$ is zero. As the interior measure is never negative and never greater than the exterior measure, it must also be zero. It follows that $E$ is measurable and has the measure zero.

From the foregoing conclusion concerning enumerable sets, it follows, as already pointed out, that the rational numbers in any closed interval form a measurable set. The set of irrational numbers in a given interval is also measurable, since they form the complementary set of one which is measurable.

The above illustration shows that of the definitions of content and measure that measure is the broader. Any point set that has a definite content has also a definite measure and the two are equal. The converse, however, is not true. If the given set is bounded and closed, it follows that the outer content and the measure are always the same[1]; for, by the Heine-Borel theorem (Theorem IV, Art. 16) the points can then be inclosed in a finite number of non-overlapping open intervals, the greatest lower bound of the sum of whose lengths is the outer content of the set.

As in the case of content, the definition of measure can be extended to point sets of two or more dimensions. The exterior measure is the greatest lower bound of the sum of the open regions containing the given set, while the interior measure is defined by means of the exterior measure of the complementary set.

The exterior measure of a set of points $E$ may be regarded as the greatest lower bound of the measure of the open sets containing $E$, while the interior measure is the least upper bound of the measure of the closed sets contained in $E$. For, as we have seen, every open set can be expressed as the sum of a finite or enumerably infinite number of open intervals. The sum of the lengths of these open intervals is the measure of the set of points defined by them. Hence, from the definition of exterior measure, it follows that the exterior measure of $E$ is the greatest lower bound of the measure of the open sets containing $E$. Likewise, the exterior measure $m_e[C(E)]$ is the greatest lower bound of the measure of the open sets containing $C(E)$. The interior measure $m_i(E)$, that is $l - m_e[C(E)]$, is then the least upper bound of the measure of the complementary sets of points, namely the closed sets contained in $E$. This discussion establishes the following theorem.

---

[1] As already stated, the outer content in the Jordan sense is identical with the term content as used by Cantor. For closed sets the term measure is identical with the term content of Cantor.

THEOREM I. *A necessary and sufficient condition that a given set $E$ is measurable is that there exists an open set $E_1$ containing $E$ and a closed set $E_2$ contained in $E$ such that*

$$E_1 > E > E_2,\ m(E_1) - m(E_2) < \epsilon,$$

*where $E_2$ may be a null-set and $\epsilon$ is an arbitrarily small positive number.*

THEOREM II. *Every bounded, open set of points is measurable.*

As we have seen, every bounded, open set is the sum of a finite or enumerably infinite set of non-overlapping, open intervals $\{\delta_n\}$. The sum of the lengths of these intervals is the measure of the given set $E$. However, each of these open intervals can be expressed as the sum of an enumerably infinite number of closed intervals having at most end-points in common. Hence, the sum of the open intervals $\{\delta_n\}$ can be expressed as the sum of an enumerably infinite number of closed intervals. The sum of the lengths of these closed intervals is equal to the sum of the intervals $\{\delta_n\}$. It follows that the open set $E$ is contained in an open set $E_1$ and contains a closed set $E_2$ such that the measures of the two sets may be made to differ by less than an arbitrarily small positive number $\epsilon$. Consequently, the given set is measurable.

THEOREM III. *Every bounded, closed set is measurable.*

The given set is complementary to a bounded, open set, which is measurable by Theorem II. Since the complementary set is measurable, the given set is also.

We may also state a necessary and sufficient condition for measurability as follows: [1]

THEOREM IV. *A necessary and sufficient condition that a point set $E$ lying in an interval $(a, b)$ has a definite measure is that*

$$E = E_1 + E_2,$$

*where $E_2$ is a closed set and $E_1$ is a point set having an exterior measure less than an arbitrarily small positive number $\epsilon$.*

This condition is necessary. As we have seen, the exterior measure of $E$ is the greatest lower bound of the measure of the open sets containing $E$, while the interior measure is the least upper bound of the measure of the closed sets contained in $E$. There exists then an open set $A$ and a closed set $E_2$ such that

$$A > E > E_2, \text{ where } m(A) - m(E_2) < \epsilon.$$

---

[1] See de la Vallée Poussin's *Intégrales de Lebesgue*, p. 23.

From the definition of a complementary set, we have

$$C(A - E_2) = C(A) + E_2.$$

Each of these sets is closed and consequently measurable. We then have from (9)

$$m[C(A - E_2)] = m[C(A)] - m(E_2).$$

Subtracting each member of this equation from $l$, the length of the interval $(a, b)$, we have

$$l - m[C(A - E_2)] = l - m[C(A)] - m(E_2).$$

whence $$m(A - E_2) = m(A) - m(E_2).$$

If we now put $E = E_1 + E_2$, where $E_1 < A - E_2$,

we have $$m_e(E_1) \leqq m(A - E_2) = m(A) - m(E_2) < \epsilon,$$

and we have the condition stated in the theorem.

The given condition is also sufficient. The given set can be written

$$E = E_1 + E_2,$$

where $E_2$ is a closed subset of $E$. We have from equation (12)

$$m_e(E) \leqq m_e(E_1) + m_e(E_2). \tag{13}$$

The set $E_2$ is closed and consequently is measurable. We then have,

$$m_e(E_2) = m(E_2). \tag{14}$$

By hypothesis, we have $$m_e(E_1) < \epsilon. \tag{15}$$

From (13) we have, by aid of (14) and (15),

$$m_e(E) \leqq m(E_2) + \epsilon. \tag{16}$$

Remembering that $m_i(E)$ is the least upper bound of the measures of the closed sets $E_2$ contained in $E$, we have from (15)

$$m_e(E) \leqq m_i(E) + \epsilon,$$

or $$m_e(E) - m_i(E) < \epsilon.$$

As the choice of $\epsilon$ is arbitrary, this relation must hold for every value of $\epsilon$ however small. It follows that $E$ satisfies the definition of measurability. This conclusion completes the proof. The theorem is equivalent to saying that the measure of $E$ is the limit of the measure of the closed sets contained in $E$.

THEOREM V. *Given a finite or enumerably infinite number of measurable point sets*

$$E_1,\ E_2,\ \cdots,\ E_n,\ \cdots$$

*contained in a finite interval* $(a, b)$; *then the sum and the product of the sets are also measurable.*[1]

We shall first consider the product. From Theorem IV we have for an arbitrarily small positive number $\epsilon$

$$\begin{aligned} E_1 &= E_{11} + E_{12}, \text{ where } m_eE_{11} < \frac{\epsilon}{2}, \\ E_2 &= E_{21} + E_{22}, \text{ where } m_eE_{21} < \frac{\epsilon}{2^2}, \\ &\cdots\cdots\cdots\cdots\cdots \\ E_n &= E_{n1} + E_{n2}, \text{ where } m_eE_{n1} < \frac{\epsilon}{2^n}. \\ &\cdots\cdots\cdots\cdots\cdots, \end{aligned} \tag{17}$$

where $E_{12}, E_{22}, \cdots, E_{n2}, \cdots$ are closed sets. Forming the product, we have

$$(E_1 \cdot E_2 \cdot \ \cdots \ \cdot E_n \cdot \ \cdots) = (E_{11} + E_{12})\,(E_{21} + E_{22}) \ \cdots \ (E_{n1} + E_{n2}) \ \cdots . \tag{18}$$

Each point in the product which is not common to all the sets $E_{n2}$ must be contained in at least one of the sets $E_{n1}$. Since

$$m_e(E_{n1}) < \frac{\epsilon}{2^n}$$

for all values of $n$, it follows that (18) can be written in the form

$$(E_1 \cdot E_2 \cdot \ \cdots \ \cdot E_n \cdot \ \cdots) = E_{12} \cdot E_{22} \cdot \ \cdots \ \cdot E_{n2} \cdot \ \cdots + A,$$

where the product $(E_{12} \cdot E_{22} \cdot \ \cdots \ \cdot E_{n2} \cdot \ \cdots)$ is a closed set and $A$ is a point set having an exterior measure which is less than $\epsilon$; for, we have

$$A < E_{11} + E_{21} + \cdots + E_{n1} + \cdots$$

and hence

$$\begin{aligned} m_e(A) &< m_e(E_{11}) + m_e(E_{21}) + \cdots + m_e(E_{n1}) + \cdots \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2^2} + \cdots + \frac{\epsilon}{2^n} + \cdots = \epsilon. \end{aligned}$$

This result is, by Theorem IV, the condition that the product is measurable.

---

[1] See de la Vallée Poussin's *Intégrales de Lebesgue*, pp. 24–25.

We shall now consider the sum

$$E = E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot \cdot.$$

The complementary set $C(E)$ contains those points found in none of the sets $E_n$; that is, it is the product of the complementary sets $C(E_n)$. We may then write

$$C(E) = C(E_1) \cdot C(E_2) \cdot \cdot \cdot \cdot \cdot C(E_n) \cdot \cdot \cdot \cdot \cdot.$$

Being a product, $C(E)$ is measurable by the foregoing discussion, and hence $E$, being the complement of $C(E)$ is also measurable.

THEOREM VI. *Given a finite or enumerably infinite set of measurable point sets contained in an interval* $(a, b)$, *no two sets having points in common; then*

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) = m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n) + \cdot \cdot \cdot \cdot.$$

We shall first consider the sum of two sets $E_1$ and $E_2$. By aid of Theorem IV, we may write

$$E_1 = A_1 + B_1, \ E_2 = A_2 + B_2,$$

where $A_1$, $A_2$ are closed sets and the exterior measures of $B_1$, $B_2$ are arbitrarily small. Hence, we have

$$E_1 + E_2 = (A_1 + A_2) + (B_1 + B_2).$$

The sets $A_1$, $A_2$ are closed and have no points in common; hence, we have from (9)

$$m(A_1 + A_2) = m(A_1) + m(A_2).$$

The measures of $E_1$, $E_2$ are respectively the least upper bounds of the measures of the closed sets $A_1$, $A_2$ contained in them. As these measures approach their least upper bounds, the exterior measure of $(B_1 + B_2)$ becomes less than any arbitrarily small positive number $\epsilon$. In the limit we have

$$m(E_1 + E_2) = m(E_1) + m(E_2). \tag{19}$$

It is at once evident that this result can be extended to the measure of the sum of any finite number of sets.

We shall now consider the sum of an enumerably infinite number of point sets. Since this sum is measurable, the measure is equal to the exterior measure and hence from (12), we have

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) \leq m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n) + \cdot \cdot \cdot \cdot. \tag{20}$$

On the other hand, we may write

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) = m(E_1 + E_2 + \cdot \cdot \cdot + E_n) + m(R_n) = m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n) + m(R_n), \quad (21)$$

where the point set

$$R_n \equiv E_{n+1} + E_{n+2} + \cdot \cdot \cdot$$

is also a measurable set by Theorem V. From (21) we have for all values of $n$

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) \geqq m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n). \quad (22)$$

As $n$ increases, the right-hand member of this inequality can not diminish since the measure of each set added is necessarily zero or some positive number. Hence,

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) \geqq m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n) + \cdot \cdot \cdot. \quad (23)$$

By comparison of the inequalities (20) and (23), we obtain

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) = m(E_1) + m(E_2) + \cdot \cdot \cdot + m(E_n) + \cdot \cdot \cdot,$$

which was to be demonstrated.

THEOREM VII. *If $E_1$, $E_2$ are measurable point sets contained in $(a, b)$ and $E_1 > E_2$, then $E_1 - E_2$ is measurable and*

$$m(E_1 - E_2) = m(E_1) - m(E_2).$$

The point set $(E_1 - E_2)$ is composed of those points of $E_1$ which are not points of $E_2$. The complementary set $C(E_1 - E_2)$ is then made up of the points of $E_2$ and those points of $(a, b)$ not contained in $E_1$, that is the points of $C(E_1)$. We may then write

$$C(E_1 - E_2) = C(E_1) + E_2,$$

and hence we have

$$m[C(E_1 - E_2)] = m[C(E_1) + E_2]. \quad (24)$$

By Theorem VI, we have

$$m[C(E_1) + E_2] = m[C(E_1)] + m(E_2). \quad (25)$$

Consequently, we obtain from (24) and (25)

$$m[C(E_1 - E_2)] = m[C(E_1)] + m(E_2).$$

Since $E_1$ is measurable, $C(E_1)$ is also measurable. Then from the foregoing equation it follows that $C(E_1 - E_2)$ is measurable and consequently $(E_1 - E_2)$ is also measurable. By the addition theorem, we may then write

$$m(E_1 - E_2) + m(E_2) = m(E_1),$$

or

$$m(E_1 - E_2) = m(E_1) - m(E_2),$$

which was to be proved.

For the limit of a sequence of point sets, we have the following theorems.

THEOREM VIII. *If we have given a sequence of measurable point sets contained in an interval* $(a, b)$ *such that*

$$E_1 < E_2 < \cdot \cdot \cdot < E_n < \cdot \cdot \cdot,$$

*then*

$$\underset{n\to\infty}{L}\, m(E_n) = m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot).$$

The sum $E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot$ is composed of the points of $E_1$ and those of the succeeding sets which do not appear in the previous sets; that is,

$$E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot = E_1 + (E_2 - E_1) + \cdot \cdot \cdot + (E_n - E_{n-1}) + \cdot \cdot \cdot .$$

The sum expressed in the right-hand member of this equation is also the limit of $E_n$ as $n$ increases indefinitely. By Theorem VI we have, therefore,

$$m(E_1 + E_2 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot) = m(E_1) + m(E_2 - E_1) + \cdot \cdot \cdot + m(E_n - E_{n-1}) + \cdot \cdot \cdot = \underset{n\to\infty}{L}\, m(E_n).$$

THEOREM IX. *If we have given a sequence of measurable sets contained in an interval* $(a, b)$ *such that*

$$E_1 > E_2 > \cdot \cdot \cdot > E_n > \cdot \cdot \cdot,$$

*then*

$$\underset{n\to\infty}{L}\, m(E_n) = m(E_1 \cdot E_2 \cdot \cdot \cdot \cdot \cdot E_n \cdot \cdot \cdot \cdot).$$

Consider the complementary set $C(E_1 \cdot E_2 \cdot \cdot \cdot)$. It includes all points of $(a, b)$ which are not common to the given sets

$$E_1, E_2, \cdot \cdot \cdot, E_n, \cdot \cdot \cdot,$$

and it therefore contains all the points which appear in the complementary set of each. We have then

$$C(E_1 \cdot E_2 \cdot \cdot \cdot \cdot \cdot E_n \cdot \cdot \cdot \cdot) = C(E_1) + C(E_2) + \cdot \cdot \cdot + C(E_n) + \cdot \cdot \cdot .$$

Hence, we have

$$m[C(E_1 \cdot E_2 \cdot \quad \cdot \cdot \cdot \quad \cdot E_n \cdot \quad \cdot \cdot \cdot)] = m[C(E_1) + C(E_2) + \cdot \cdot \cdot + C(E_n) + \cdot \cdot \cdot]. \quad (26)$$

Moreover, we have

$$C(E_1) < C(E_2) < \cdot \cdot \cdot < C(E_n) < \cdot \cdot \cdot \cdot .$$

Since $E_n$ is measurable the complementary set $C(E_n)$ is also measurable. From Theorem VIII, we have

$$m[C(E_1) + C(E_2) + \cdot \cdot \cdot + C(E_n) + \cdot \cdot \cdot] = \underset{n \to \infty}{L}\, m[C(E_n)]. \quad (27)$$

From (26) and (27), it follows that

$$\underset{n \to \infty}{L}\, m[C(E_n)] = m[C(E_1 \cdot E_2 \cdot \quad \cdot \cdot \cdot \quad \cdot E_n \cdot \quad \cdot \cdot \cdot)].$$

Subtracting each side of this equation from the length of the interval containing the given sets, namely $b - a$, we have by aid of (7)

$$\underset{n \to \infty}{L}\, m(E_n) = m(E_1 \cdot E_2 \cdot \quad \cdot \cdot \cdot \quad \cdot E_n \cdot \quad \cdot \cdot \cdot),$$

which is the required result.

Finally, we can establish the following more general theorem.

THEOREM X. *Given the sequence of measurable point sets* $E_1, E_2, \cdot \cdot \cdot, E_n, \cdot \cdot \cdot$ *contained in an interval* $(a, b)$ *and having the limiting point set* $E$. *Then* $E$ *is measurable and we have*

$$\underset{n \to \infty}{L}\, m(E_n) = m(E).$$

In Art. 13 we have seen that the complete limit of $E_n$ of such a sequence of point sets can be expressed as the limit of the sum

$$E_n + E_{n+1} + \cdot \cdot \cdot ,$$

while the restricted limit is the limit of the product

$$(E_n \cdot E_{n+1} \cdot \quad \cdot \cdot \cdot).$$

As the given sequence has an unique limit, its complete and restricted limits must be the same and equal to its limiting set $E$. We have then from equations (1) and (2) of Art. 13

$$E = \underset{n \to \infty}{L}\, E_n = [(E_1 + E_2 + \cdot \cdot \cdot)\,(E_2 + E_3 + \cdot \cdot \cdot) \cdot \cdot \cdot], \quad (28)$$

$$E = \underset{n \to \infty}{L}\, E_n = [(E_1 \cdot E_2 \cdot \quad \cdot \cdot \cdot) + (E_2 \cdot E_3 \cdot \quad \cdot \cdot \cdot) + \cdot \cdot \cdot]. \quad (29)$$

The last member of the first of these relations is the product of a sequence of decreasing point sets, and hence by Theorem IX, we have

$$\underset{n \to \infty}{L}\, m(E_n + E_{n+1} + \cdot \cdot \cdot) = m[(E_1 + E_2 + \cdot \cdot \cdot)\,(E_2 + E_3 + \cdot \cdot \cdot)\,(E_3 + E_4 + \cdot \cdot \cdot) \cdot \cdot \cdot]. \quad (30)$$

From (28) we obtain

$$m(E) = m[(E_1 + E_2 + \cdot\ \cdot\ \cdot)(E_2 + E_3 + \cdot\ \cdot\ \cdot) \cdot\ \cdot\ \cdot],$$

and by comparison with (30), we have

$$m(E) = \underset{n\to\infty}{L}\, m(E_n + E_{n+1} + \cdot\ \cdot\ \cdot). \tag{31}$$

But

$$(E_n + E_{n+1} + \cdot\ \cdot\ \cdot) \geqq E_n,$$

and consequently from (31) we obtain

$$m(E) \geqq \underset{n\to\infty}{L}\, m(E_n).$$

In a similar manner we obtain from (29), by aid of Theorem VIII,

$$m(E) \leqq \underset{n\to\infty}{L}\, m(E_n).$$

Consequently, we must have

$$m(E) = \underset{n\to\infty}{L}\, m(E_n),$$

as stated in the theorem.

## EXERCISES

**1.** Can an isolated set be everywhere dense? be non-enumerable? perfect? have the measure zero? be dense in itself? have the cardinal number of the continuum? have points of condensation?

**2.** Can a set which is no subinterval everywhere dense have an outer content greater than zero? Illustrate your conclusion. [See *Proceedings of London Math. Soc.*, Vol. 6, p. 148.]

**3.** Give a set, other than the continuum, having $\frac{1}{2}$ as a limiting point. Is the set enumerable or non-enumerable? closed or open? Could it be neither closed nor open? Pick out the inner points, the boundary points. Are the latter identical with the limiting points of the set?

**4.** Can a set have only inner points? only boundary points? Illustrate. What relation do the boundary points of the set have to the complementary set? Can the boundary points of a set be also points of condensation?

**5.** Is every set of the first category enumerable? Illustrate.

**6.** Let $E'$ be a discrete set. Is $E$ also discrete? everywhere dense? Must $E$ be closed? Give reasons for your conclusions.

**7.** Show that the sum $E$ of a finite or enumerably infinite number of enumerable sets is a set of the first category. Can it be a perfect set? Can it be everywhere dense? What can be said of the measure of $E$? Can it be closed? What statements can be made of the set $C(E)$?

**8.** An **algebraic number** of rank $n$ is defined as a root of an irreducible algebraic equation of degree $n$, with rational coefficients, which is not a root of such an equation of lower degree. Show that the set $E$ of all algebraic numbers is enumerable. Does this set $E$ include all rational numbers? Does it contain any irrational numbers? Is $E$ of the first or the second category? [See Young, *Theory of Point Sets*, p. 44.]

**9.** All real numbers that are not algebraic are called transcendental. Do such numbers form an enumerable set? How does the cardinal number of this set compare with that of the set of rational numbers? the irrational numbers? the continuum? Show that the transcendental numbers contained in the interval (0, 1) form a set of measure one.

**10.** Can a perfect set be enumerable? isolated? consist of boundary points only? have the measure zero? Give your reason in each case and illustrate your conclusion.

**11.** Could one obtain an enumerable set by suppressing from the continuum an enumerable set an indefinitely large number of times? Give your reason for the conclusion. Could one in this same way obtain a set of measure zero?

**12.** If the derived set $E'$ is enumerable, could one exhaust $E$ by the continued separation of isolated subsets? Illustrate your conclusion.

**13.** In case a set $E$ is dense in itself but not closed, show that $E'$ can not be enumerable but must have the cardinal number of the continuum.

**14.** Could a closed set be enumerable? have a complementary set having the cardinal number of the continuum? have the measure zero? Illustrate each conclusion.

**15.** Can a set of the first species be everywhere dense in a given interval?

**16.** Do the irrational points form a closed set? an open set? What points are boundary points of the set? Does it have a closed subset? Does it contain points of condensation? Reconcile your conclusions with the statements of Theorem I, Art. 16, and Theorem III, Art. 17.

**17.** Given the closed interval (0, 1) and let $\lambda$ satisfy the condition $0 \leqq \lambda \leqq 1$. Drop from the interval (0, 1) and also in succession from the remaining free subintervals a portion of length $l_n$ ($n = 1, 2, 3 \ldots$) satisfying the general condition that

$$l_1 + 2l_2 + 2^2 l_3 + \cdots + 2^{n-1} l_n = \lambda - \frac{\lambda}{n+2}.$$

Let $E$ be the set of end-points of the intervals $\{l_n\}$ plus their limiting points. Show that $E$ is a perfect set and by the proper choice of $\lambda$ may have any desired content from 0 to 1.

**18.** Is the sum of a finite number of discrete sets necessarily discrete? Can such a sum be everywhere dense in a given interval? Illustrate.

## CHAPTER III

# CONTINUITY AND DISCONTINUITY OF FUNCTIONS

**22. Fundamental definitions.** If a symbol may be assigned different numerical values in the course of a discussion, it is spoken of as a **variable**. The aggregate of real numbers constitutes the domain of a real variable. The results of any arithmetical operation, other than the extraction of even roots of negative values, on a real variable is again a real variable. To the various values which such a variable may assume correspond the points on a straight line. We say that the variable $x$ lies in the interval $(a, b)$ if the values which $x$ may take lie between $a$ and $b$. If $a$ and $b$ are included in the interval, it is said to be a **closed interval** and we write

$$a \leqq x \leqq b.$$

If $a$ and $b$ are not included, we say that $x$ lies in the **open interval** $(a, b)$ and write

$$a < x < b.$$

We shall regard a given interval as closed unless otherwise stated.

The elementary notion of a function dates from the seventeenth century,[1] and its early use was closely associated with the investigations of Descartes and Fermat in the applications of algebra to the solution of geometric problems. Leibniz was the first to use the term function in a sense approaching its present use, and John Bernoulli and Euler were the first to introduce anything like a precise definition. The definition which is now commonly accepted by mathematicians is due to Dirichlet and may be stated as follows.

Having given two variables $x$ and $y$, $y$ is said to be a **function** of $x$ if for each given value of $x$ there is determined a definite value or set of values of $y$. We write $y = f(x)$, and speak of $x$ as the **independent variable** and of $y$ as the **dependent variable.**

It will be observed that the independent variable $x$ need not take all values of the real continuum, and in fact a function $f(x)$ can just as well be defined for a set of values $\{x\}$ as for the continuum. What is essential in the definition is that for every value which $x$ does take there is uniquely determined one or more definite values of the function $f(x)$. If for any value of $x$ the function fails to have a definite value, then the function can not be said to be defined for that value. For example,

---

[1] See *Encyclop. des sci. math.* II, 1, pp. 1–6.

$$y = \sin \frac{1}{x}$$

is not defined for $x = 0$; for, when $x$ takes the value zero the value of $y$ is not definitely determined by this relation. If we assign to $y$ a definite value for $x = 0$, say by putting

$$f(0) = 0,$$

then $f(x)$ is fully defined for all values of $x$.

A function $f(x)$ may be defined for each value of $x$ which enters into consideration, although one may not care or may not be able to write the function as an analytic expression of $x$. In other words, the functional relation may be stated otherwise than by means of an analytic formula. For example, $f(x)$ is fully defined in the interval $(0, 1)$ if we say that for all rational values of $x$ in this interval it takes the value one and for the irrational values of $x$ it has the value zero.

The definition of a function may be generalized to include functions of two or more independent variables. Thus, we say that

$$z = f(x, y)$$

is a function of the two independent variables $x$ and $y$, if for each pair of values of $x$ and $y$ of a certain set $\{(x, y)\}$ over which the point $(x, y)$ ranges there is determined a definite value or set of values of $z$.

The definition of a function as formulated leads to some general properties of the function itself.[1] In order to be made the basis of analysis, however, we shall need to know certain other properties of the function such as continuity, differentiability, etc., to be discussed later.

If to each value of the independent variable there corresponds but one value of the function, then the function is said to be **single-valued.** If, on the other hand, two or more values of the function correspond to some or all of the values assigned to the independent variable, the function is called **multiple-valued.** In the discussions to follow, $f(x)$ will be regarded as single-valued unless otherwise stated.

---

[1] Cf. Blumberg, *New Properties of all Real Functions*, Trans. Amer. Math. Soc. (1922), Vol. 24, pp. 113–128.

A **rational integral function** or polynomial is a function of the form

$$a_0x^n + a_1x^{n-1} + \cdot \quad \cdot \quad \cdot + a_{n-1}x + a_n,$$

where $a_0, a_1, \cdot \cdot \cdot , a_n$ are constants and $n$ is a positive integer.

By a **rational function** is understood the quotient of one rational integral function by another. Thus, if $f(x)$ and $\phi(x)$ are rational integral functions and $\phi(x)$ is not identically zero, then

$$F(x) = \frac{f(x)}{\phi(x)}$$

is a rational function.

If the relation between $x$ and $y$ is given by means of an equation of the form

$$y^n + f_1(x)y^{n-1} + f_2(x)y^{n-2} + \cdot \quad \cdot \quad \cdot + f_n(x) = 0,$$

where $f_1(x), f_2(x), \cdot \cdot \cdot , f_n(x)$ are rational functions of $x$, then $y$ is said to be an **algebraic function** of $x$.

Functions which are not algebraic are called **transcendental functions.** The trigonometric functions, the exponential function, and the logarithmic function are illustrations of such functions.

As we have seen, a functional relation between two variables is nothing else than an established correspondence between two sets of values. As $x$ takes the values of one set, $y = f(x)$ takes the corresponding values of the other set, which may be regarded as a point set $\{y\}$ on the $Y$-axis. If the set $\{y\}$ is bounded, then the function $f(x)$ is said to be **bounded** on the set $\{x\}$. The least upper bound of the set $\{y\}$ is spoken of as the **least upper bound** of the function. Likewise, the upper limit of $\{y\}$ is called the **upper limit** of the function. From the discussion in Art. 11, it will be seen that the two values may or may not be the same; that is, the least upper bound may be either the upper limit or an extreme value distinct from the upper limit. In case the least upper bound is a point of the set $\{y\}$ it is called the **maximum value** of the function. Similarly, we can formulate the corresponding definitions for the greatest lower bound, lower limit, and minimum value of a function.

We shall need later to consider the least upper bound and the upper limit of a sequence of functions $\{f_n(x)\}$. For the purpose of definition, we shall make use of an auxiliary function $\phi(x)$ defined as follows. For each particular value of $x$ in the given interval, let $\phi(x)$ be the least upper bound of the corresponding set of values

$\{f_n(x)\}$, disregarding the order in which the functions $f_n(x)$ may occur. The function $\phi(x)$ thus defined for all values of $x$ in the given interval is said to be the **upper boundary function** of the given sequence of functions $\{f_n(x)\}$. A similar definition may be formulated for the **lower boundary function** of the sequence.

Denote by $f(x)$ the function defined by the sequence $\{f_n(x)\}$. If the sequence converges for a particular value of $x$, then the corresponding value of the limiting function $f(x)$ is uniquely determined. For any value $x = x_0$, the sequence may fail to converge either by becoming definitely infinite or by oscillating between finite or infinite bounds. In case of oscillation, one may, by a proper choice of the values of $n$, make $f_n(x_0)$ converge to any one of a closed set of values. The aggregate of the least upper bounds of these sets as $x$ takes all values in the given interval is called the **upper limiting function** defined by $\{f_n(x)\}$. In a similar manner the **lower limiting function** may be defined. In case the upper limiting function and the lower limiting function are the same, then the sequence has a unique limiting function.

**23. Definition of a limit.** If a variable $x$ is defined for a linear continuum and takes all values in the neighborhood of a definite number $x_0$, we say that $x$ has the limit $x_0$ or **approaches** the limit $x_0$. This use of the term approach does not imply, however, any particular order or succession in which these values are taken. It merely asserts that $x$ may be assigned at pleasure any value in the neighborhood, deleted or not, of $x_0$. If the variable $x$ is given for only a set of values $E$, then it is said to approach a definite value $x_0$ if it takes all the values of $E$ in the neighborhood of $x_0$, provided of course that those values are dense at $x_0$. It is usual to indicate the fact that $x$ approaches $x_0$ by writing $x \to x_0$, or by $x \doteq x_0$.

Corresponding to the values of $x$ in the neighborhood of $x_0$ the function $f(x)$ takes a set of values which may also approach a definite limit $A$ as $x$ approaches $x_0$. The **limit of a function** may then be defined as follows. Having given an arbitrarily small positive number $\epsilon$, if there exists a positive number $\delta$ such that

$$|f(x) - A| < \epsilon \tag{1}$$

for all values of $x \neq x_0$ which satisfy the condition $|x - x_0| < \delta$, then $A$ is said to be the limit of $f(x)$ as $x$ approaches $x_0$. Symbolically we also indicate this relation by writing

$$\underset{x \to x_0}{L} f(x) = A.$$

If the limit exists, it follows that we obtain the same limiting value if $x$ is allowed to approach its limit through any particular set of values dense at $x_0$ as when it takes all values in the neighborhood of this point. If $f(x)$ is defined for a set $E$ rather than for the continuum, then $x_0$ must be a limiting point of the set and $x$ must take all values of $E$ in the neighborhood of $x_0$.

The existence of the limit does not depend upon the value of the function at the limiting point itself but depends entirely upon the values in its deleted neighborhood; that is, it is immaterial so far as the mere existence of the limit is concerned whether the limiting value $A$ has any relation to the value of $f(x)$ for $x = x_0$ or in fact whether $f(x)$ is defined for this value of $x$. It is essential to observe the order in which $\epsilon$ and $\delta$ are chosen, namely that $\epsilon$ is chosen first while $\delta$ is determined by that choice. The value of $\delta$ therefore depends upon $\epsilon$ and $x_0$.

The geometric significance of a limit may be shown as follows. Upon the perpendicular through the limiting point $x = x_0$ locate the point $y = A$. On this same perpendicular lay off $A \pm \epsilon$. Corresponding to the value of $\epsilon$ there is determined a positive number $\delta$, and the existence of the inequality (1) means geometrically that for every value of $x$ in the interval $(x_0 - \delta, x_0 + \delta)$ the corresponding value of $f(x)$ lies within the rectangle bounded by the lines

$$y = A \pm \epsilon,\ x = x_0 \pm \delta.$$

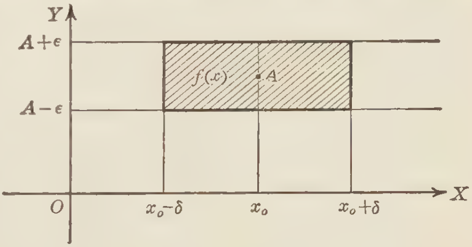


FIG. 10.

The points representing $f(x)$ are dense in the two-dimensional neighborhood of $A$. The value of $\epsilon$ is arbitrarily small, and if its value is changed the value of $\delta$ may change, but the existence of the limit is equivalent to saying that for every value assigned to $\epsilon$ there exists a rectangle of area different from zero satisfying the foregoing conditions.

**Ex. 1.** Given the function $f(x) = \sin \dfrac{1}{x}$ for $x \neq 0$. Consider the limit as $x$ approaches 0.

This function is not defined for $x = 0$, but, as we have seen, the existence of a limit depends on the values of the variable in the deleted neighborhood of the point and not on the value at the point. In this case, however, the limit does not exist as no value $A$ can be found such that the inequality (1) is satisfied; for, in every neighborhood of $x = 0$ the value of the function takes all values between $+1$ and $-1$. This is also evident from the graph of the function, which is given in Fig. 11.

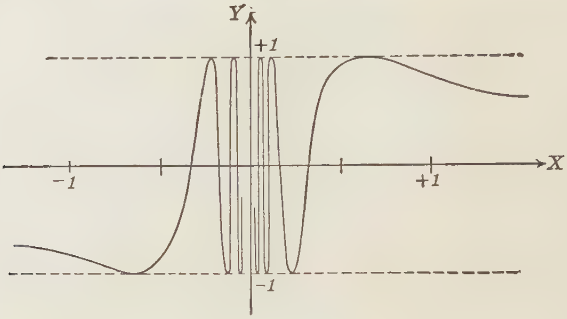


FIG. 11.

**Ex. 2.** Consider the limit as $x$ approaches 0 of the function $y = x \sin \frac{1}{x}$, $x \neq 0$. The oscillations of the function in this case lie between the lines $y = \pm x$. The graph is given in Fig. 12.

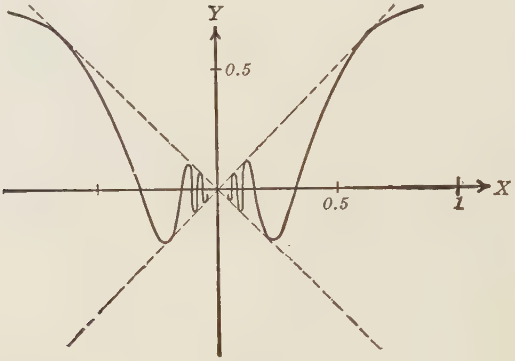


FIG. 12.

We have
$$\underset{n\to\infty}{L}\, x \sin \frac{1}{x} = 0;$$
for, assigning any arbitrarily small value to $\epsilon$, there exists a $\delta > 0$ such that
$$\left| \delta \sin \frac{1}{\delta} - 0 \right| = \left| \delta \sin \frac{1}{\delta} \right| < \epsilon,$$
since the sine is at most equal to unity and $\delta$ can be taken sufficiently small to make the product less than $\epsilon$.

In Ex. 1 the limit fails to exist because in every neighborhood of the limiting point the function has an infinite number of oscillations each of which is greater than a fixed positive number. The limit may also fail to exist for other reasons; for example, the limit of $f(x)$ for values of $x$ on one side of $x_0$ may be different from the limit for values of $x$ on the other side. The following is such a case.

**Ex. 3.** Consider the limit as $x$ approaches $\frac{\pi}{2}$ of $f(x) = \frac{e^{\tan x} - 1}{e^{\tan x} + 1}$.

If we restrict $x$ to values to the left of $\frac{\pi}{2}$, we have, by aid of Theorem I, Art. 24,

$$\underset{x \to \frac{\pi}{2}}{L} \frac{e^{\tan x} - 1}{e^{\tan x} + 1} = \underset{x \to \frac{\pi}{2}}{L} \frac{1 - \frac{1}{e^{\tan x}}}{1 + \frac{1}{e^{\tan x}}} = \frac{1 - 0}{1 + 0} = 1.$$

If we consider only values to the right of $\frac{\pi}{2}$, we have

$$\underset{x \to \frac{\pi}{2}}{L} \frac{e^{\tan x} - 1}{e^{\tan x} + 1} = \frac{0 - 1}{0 + 1} = -1.$$

Hence, the limit $\underset{x \to \frac{\pi}{2}}{L} f(x)$ can not be said to exist.

If, as in the foregoing example, the function has a limit for values of the independent variable lying to the right of the critical point and to the left of it, it is often convenient to speak of these limits as the **right-hand** and **left-hand limits**; even though they may be different. We may indicate symbolically the right-hand and the left-hand limits at any point $x_0$ by writing respectively

$$\underset{x \to x_0+}{L} f(x), \quad \underset{x \to x_0-}{L} f(x).$$

If the function increases without limit, that is if it becomes infinite, as the variable approaches the critical point, then of course the limit does not exist, since there is no definite value $A$ which satisfies the inequality (1).

If we have given a function $f(x, y)$ of two independent variables, limits of a kind different from those already discussed come into consideration. The variables may approach a limiting position either by varying separately or simultaneously. If $(x_0, y_0)$ is the limiting point, we have then the limits

$$\underset{x \to x_0}{L} \underset{y \to y_0}{L} f(x, y), \quad \underset{y \to y_0}{L} \underset{x \to x_0}{L} f(x, y), \quad \underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y),$$

the last indicating that $x$ and $y$ approach their limiting values simultaneously. We shall speak of the first two types as iterated double limits or simply **iterated limits** and of the last as a simultaneous double limit or more briefly as a **simultaneous limit.** An iterated limit presents nothing essentially new, as it is merely a limit of a limit.[1] However, in simultaneous limits we have something quite different from the ordinary simple limit.

We say that the simultaneous limit $\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x, y)$ exists and is equal to $A$, if for an arbitrarily small positive number $\epsilon$ there exists a positive number $\delta$ such that we have

$$|f(x, y) - A| < \epsilon, \tag{2}$$

for all values of $x$ and $y$ in the neighborhood of $(x_0, y_0)$ defined by the inequalities

$$|x - x_0| < \delta, \ |y - y_0| < \delta,$$

where the point $(x_0, y_0)$ itself is omitted from consideration.

As there are two independent variables, the geometric interpretation is in three-dimensional space. It means that on the perpendicular to the $(x, y)$-plane at $(x_0, y_0)$ there exists a point $A$ such that, for an arbitrarily small $\epsilon$, there is in the $(x, y)$-plane a square whose sides are

$$x = x_0 \pm \delta, \ y = y_0 \pm \delta,$$

such that for every point $(x, y)$ other than $(x_0, y_0)$ within this square the corresponding value of $f(x, y)$ lies within a rectangular parallelopiped having $A$ as a center and bounded by the planes whose equations are

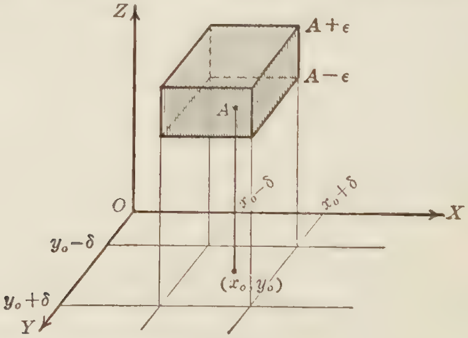


FIG. 13.

$$x = x_0 \pm \delta, \ y = y_0 \pm \delta, \ z = A \pm \epsilon.$$

---

[1] Some writers take the position that the iterated limit $\underset{x\to x_0}{L}\ \underset{y\to y_0}{L} f(x, y)$ may exist even though the limit $\underset{y\to y_0}{L} f(x, y)$ does not exist. Cf. Hobson, *Theory of Functions of a Real Variable*, Vol. 2, p. 383.

The existence of a simultaneous limit carries with it a wider range of consequences than is the case with single limits. For example, not only must we have the same limiting value if the variable point $(x, y)$ approaches the limiting point $(x_0, y_0)$ through any set of values dense at that point, but we must also have the same limiting value as the variable point approaches its limiting position along any curve whatsoever. Consequently, if we can find two methods of approach to the limiting point which give different limiting values, that is evidence that the simultaneous limit does not exist. This fact can often be used in determining the non-existence of a simultaneous limit.

**Ex. 4.** Consider the simultaneous limit of $f(x, y) = \dfrac{xy}{x^2 + y^2}$ as $x \rightarrow 0$, $y \rightarrow 0$.

If the variables be made to approach the origin along the line $y = x$, we have upon putting $y = x$

$$\underset{x \rightarrow 0}{L} \frac{x^2}{x^2 + x^2} = \frac{1}{2}.$$

If, however, we first allow $x$ to approach zero along a line parallel to the $X$-axis and then allow $y$ to approach zero along the $Y$-axis, we have

$$\underset{y \rightarrow 0}{L} \; \underset{x \rightarrow 0}{L} \frac{xy}{x^2 + y^2} = \underset{y \rightarrow 0}{L} \frac{0}{0 + y^2} = 0.$$

The different results obtained by these two methods of approach to the limiting point show that the simultaneous limit does not exist.

If a simultaneous limit does exist, it follows that the limiting value obtained by assuming any convenient relation between the variables and allowing them to approach the limiting position along the corresponding curve gives the value of the limit. The converse is not true. For example, the same limiting value of the function may be obtained as the variables are allowed to approach their limiting values along all straight lines passing through the limiting point and yet the simultaneous limit need not exist. The following is an illustration.

**Ex. 5.** Consider the simultaneous limit of $\dfrac{xy^3}{x^2 + y^2}$ as $x \rightarrow 0$, $y \rightarrow 0$.

If we put $y = mx$, which is the general equation of a line through the origin and then allow $x$ to approach zero we have

$$\underset{x \rightarrow 0}{L} \frac{m^3 x^4}{x^2 + m^6 x^6} = \underset{x \rightarrow 0}{L} \frac{m_3 x^2}{1 + m^6 x^4} = 0.$$

However, if we put $x = y^3$ and then allow $y$ to approach zero, we obtain

$$\underset{y \rightarrow 0}{L} \frac{y^6}{y^6 + y^6} = \frac{1}{2}.$$

The simultaneous limit, therefore, does not exist; for, that requires that we obtain the same limiting value by all possible methods of approach to the limiting point.

If the simultaneous limit $\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x, y)$ exists, then the single limits $\underset{x\to x_0}{L} f(x, y_0)$, $\underset{y\to y_0}{L} f(x_0, y)$ also exist. It does not follow, however, that the single limits $\underset{x\to x_0}{L} f(x, y)$, $\underset{y\to y_0}{L} f(x, y)$ exist for $y \neq y_0$, $x \neq x_0$, respectively, even for constant values of $y$ and $x$ differing ever so little from $y_0$ and $x_0$. This statement is verified by the following illustration.

**Ex. 6.** Consider the simultaneous limit of $y \sin \frac{1}{x}$ for $x = 0$, $y = 0$.

We have
$$\underset{\substack{x\to 0\\ y\to 0}}{L} y \sin \frac{1}{x} = 0;$$
for, we have $\left|0 - y \sin \frac{1}{x}\right| = \left|y \sin \frac{1}{x}\right| < \epsilon$, $0 < |x| < \delta$, $|y| < \delta$.

However, for any constant value of $y \neq 0$, say $y = y_1$, we have
$$\underset{x\to 0}{L}\, y_1 \sin \frac{1}{x} = y_1 \underset{x\to 0}{L} \sin \frac{1}{x}.$$
But the limit $\underset{x\to 0}{L} \sin \frac{1}{x}$ does not exist, as we have seen (Ex. 1).

The inequalities (1) and (2) may be expressed in an equivalent form by means of the equations
$$f(x) = A \pm \epsilon(x),$$
$$f(x, y) = A \pm \epsilon(x, y),$$
where $\epsilon$ may be made as small as we please by the proper choice of $x$ or in the second case of the independent variables $x$ and $y$.

**24. Laws of operation with limits.** From the study of elementary mathematics the reader is already familiar with many of the laws of operation with limits. The elementary laws of algebra apply so long as only a finite number of functions are involved. If we have an infinite number of functions it no longer necessarily follows that the sum, product, etc., of the limits of the separate functions is the limit of the sum, product, etc., of the separate functions. In such cases, as we shall see later, it is necessary to impose certain conditions upon the functions themselves and to study more carefully the manner of the convergence. There are certain other laws of operation with limits which we shall now consider.

THEOREM I. *If* $\underset{y\to b}{L} F(y) = F(b)$, *where* $y$ *is such a function of* $x$ *that* $\underset{x\to a}{L} y(x) = b$, *then* $\underset{x\to a}{L} F\{y(x)\} = F\{\underset{x\to a}{L} y(x)\}$.

By hypothesis $$\underset{y\to b}{L}\, F(y) = F(b). \tag{1}$$

Since $y$ is a function of $x$, we have by substituting this function for $y$ in $F(y)$ another function of $x$, say $f(x) \equiv F\{y(x)\}$. We also have the condition that $y$ approaches $b$ as $x$ approaches $a$. Hence, we have

$$\underset{x\to a}{L}\, F\{y(x)\} \equiv \underset{x\to a}{L}\, f(x) = \underset{y\to b}{L}\, F(y). \tag{2}$$

By the conditions given in the theorem, we also have

$$F(b) = F\{\underset{x\to a}{L}\, y(x)\}. \tag{3}$$

Hence, from (1), (2), (3) we have the desired result, namely

$$\underset{x\to a}{L}\, F\{y(x)\} = F\{\underset{x\to a}{L}\, y(x)\}.$$

The foregoing theorem is of value in evaluating the limits of certain functions such as the logarithmic, exponential, and trigonometric functions.

**Ex. 1.** Consider the limit $\underset{x\to 2}{L} \log \dfrac{x^2+4}{x+2}$.

From the foregoing theorem the limit of the logarithm is the logarithm of the limit, provided the limit of the logarithm is finite and determinate at the point in question. Hence, in this case

$$\underset{x\to 2}{L} \log \frac{x^2+4}{x+2} = \log \underset{x\to 2}{L} \frac{x^2+4}{x+2} = \log 2.$$

THEOREM II. *If $f(x)$ and $\phi(x)$ are two functions whose limits exist, then*

$$\underset{x\to a}{L}\, \{f(x)\}^{\phi(x)} = \{\underset{x\to a}{L}\, f(x)\}^{\underset{x\to a}{L}\,\phi(x)},$$

*provided the latter limit is not indeterminate.*

Under the given conditions, we may write

$$\{f(x)\}^{\phi(x)} = e^{\phi(x)\log f(x)}.$$

From Theorem I, we have

$$\begin{aligned}\underset{x\to a}{L}\, \{f(x)\}^{\phi(x)} = \underset{x\to a}{L}\, e^{\phi(x)\log f(x)} &= e^{\underset{x\to a}{L}\,\phi(x)\log f(x)}\\ &= e^{\underset{x\to a}{L}\,\phi(x)\log \underset{x\to a}{L}\, f(x)}\\ &= \{\underset{x\to a}{L}\, f(x)\}^{\underset{x\to a}{L}\,\phi(x)}\end{aligned}$$

provided this result does not assume an indeterminate form.

THEOREM III. *A single-valued, bounded, monotone function has a left-hand (right-hand) limit at each point of its interval of definition.*

We shall demonstrate the theorem for the case of a monotone increasing function. The proof for a monotone decreasing function is similar and can be readily supplied.

To each $x$ there corresponds one and only one value of $y$, since $f(x)$ is single-valued. The set $\{y\}$ is bounded and hence has a least upper bound $Q$. As $x$ increases, there exists to the left of each point $x_0$ a deleted neighborhood $x_0 - \delta \leqq x < x_0$, for which $f(x)$ either remains constant or increases. If $f(x)$ remains constant, it follows at once that the required limit exists. If $f(x)$ is not a constant, then for every value of $\delta > 0$ there corresponds to the values of $x$ in the foregoing neighborhood of $x_0$ an infinite set of increasing values of $y$ which approach this least upper bound as $x$ approaches $x_0$. Since the function is monotone increasing, it follows that, in this case also, the left-hand limit $\underset{x \to x_0}{L} f(x)$ must exist.

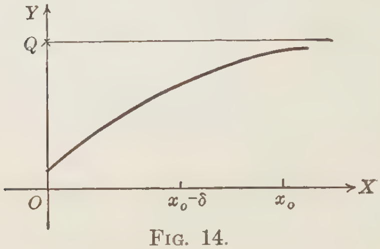


FIG. 14.

If $x$ ranges over an infinite interval, it can be shown in a similar manner that the limit $\underset{x \to \infty}{L} f(x)$ also exists. It may be shown that the right-hand limit at $x_0$ also exists. It does not follow, however, that the two limits are equal. In fact, one or both of these limits may differ from $f(x_0)$.

Another theorem having important applications is the following.

THEOREM IV. *If $\psi(x)$ and $\phi(x)$ have the same limiting value $A$ as $x$ approaches $a$, and if for all values of $x$ in a sufficiently small neighborhood of $a$, we have*

$$\psi(x) \leqq f(x) \leqq \phi(x),$$

*then it follows that*

$$\underset{x \to a}{L} f(x) = A.$$

Since the limits $\underset{x \to a}{L} \psi(x)$, $\underset{x \to a}{L} \phi(x)$ exist and are equal to $A$, we have for $x$ sufficiently near $a$

$$\psi(x) = A \pm \epsilon_1,$$
$$\phi(x) = A \pm \epsilon_2,$$

where $\epsilon_1$, $\epsilon_2$ are arbitrarily small positive numbers. Hence, for values of $x$ sufficiently near $a$, we have

$$A \pm \epsilon_1 \leqq f(x) \leqq A \pm \epsilon_2,$$

or

$$\epsilon_1 \leqq |\, f(x) - A \,| \leqq \epsilon_2.$$

Since this inequality holds for all values of $x$ sufficiently near $a$, say for $x$ such that $|\, x - a \,| < \delta$, it follows that

$$|\, f(x) - A \,| < \epsilon \text{ for } |\, x - a \,| < \delta,$$

where $\epsilon$ is an arbitrarily small positive number, and hence

$$\underset{x \to a}{L}\, f(x) = A.$$

**Ex. 2.** Prove that $\underset{x \to 0}{L}\, \frac{\sin x}{x} = 1.$

In the neighborhood of $x = 0$, we have

$$\sin x \leqq x \leqq \tan x,$$

or

$$1 \leqq \frac{x}{\sin x} \leqq \frac{1}{\cos x}.$$

But, we have

$$\underset{x \to 0}{L}\, \frac{1}{\cos x} = 1.$$

Hence, from the theorem, we have

$$\underset{x \to 0}{L}\, \frac{x}{\sin x} = \underset{x \to 0}{L}\, \frac{\sin x}{x} = 1.$$

The foregoing theorems have been established for the case of single limits. They may be extended to the case of simultaneous limits. The method of proof is similar to that for single limits and may be left to the reader to supply. There are also some theorems which are concerned primarily with simultaneous limits and iterated limits. The following is such a one.

THEOREM V. *If the simultaneous limit* $\underset{\substack{x \to x_0 \\ y \to y_0}}{L}\, f(x, y)$ *exists and is equal to* $A$ *and if the limit* $\underset{x \to x_0}{L}\, f(x, y)$ *exists for each constant value of* $y$ *in the neighborhood of* $y_0$ *and likewise the limit* $\underset{y \to y_0}{L}\, f(x, y)$ *exists for each constant value of* $x$ *in the neighborhood of* $x_0$*, then*

$$\underset{x \to x_0}{L}\ \underset{y \to y_0}{L}\, f(x, y) = \underset{y \to y_0}{L}\ \underset{x \to x_0}{L}\, f(x, y) = A.$$

By hypothesis the limit $\underset{x \to x_0}{L}\, f(x, y)$ exists for each value of $y$ in the neighborhood of $y = y_0$. The aggregate of these limiting values defines a function $F(x_0, y)$ of $y$. We may then write

$$\underset{x \to x_0}{L}\, f(x, y) = F(x_0, y), \tag{4}$$

where $F(x_0, y)$ may or may not be identical with $f(x_0, y)$. Stated as an inequality, this relation becomes, for each value of $y$ in a neighborhood of $y_0$, say in the neighborhood defined by the relation $|y-y_0| < \delta_0$,

$$| F(x_0, y) - f(x, y) | < \epsilon_1, \ | x - x_0 | < \delta_1. \tag{5}$$

From the existence of the simultaneous limit at $(x_0, y_0)$ we have

$$| f(x, y) - A | < \epsilon_2, \ | x - x_0 | < \delta_2, \ | y - y_0 | < \delta_2. \tag{6}$$

Combining the inequalities (5) and (6), we have

$$| F(x_0, y) - A | < \epsilon, \ | y - y_0 | < \delta,$$

where $\delta$ is the smaller of the two numbers $\delta_0$, $\delta_2$ and

$$\epsilon = \epsilon_1 + \epsilon_2,$$

so that $\epsilon$ is arbitrarily small. This inequality is equivalent to the relation

$$\underset{y \to y_0}{L} F(x_0, y) = A.$$

Replacing $F(x_0, y)$ by its equal given in (4), we have

$$\underset{y \to y_0}{L} \ \underset{x \to x_0}{L} f(x, y) = A.$$

In the same way, we may show that

$$\underset{x \to x_0}{L} \ \underset{y \to y_0}{L} f(x, y) = A,$$

and hence we have finally

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y) = \underset{x \to x_0}{L} \ \underset{y \to y_0}{L} f(x, y) = \underset{y \to y_0}{L} \ \underset{x \to x_0}{L} f(x, y) = A.$$

This theorem states a sufficient but not a necessary condition for the interchange of the order of iterated limits; that is, there are cases where the order may be interchanged and yet the conditions set forth in the theorem are not satisfied. The following exercise gives an illustration of such a case.

**Ex. 3.** Consider the iterated limits of $\dfrac{xy}{x^2 + y^2}$ for $x = 0$, $y = 0$.

As we have seen (Ex. 4, Art. 23) the simultaneous limit $\underset{\substack{x \to 0 \\ y \to 0}}{L} \dfrac{xy}{x^2 + y^2}$ does not exist. However, we have

$$\underset{x \to 0}{L} \ \underset{y \to 0}{L} \frac{xy}{x^2 + y^2} = \underset{x \to 0}{L} \frac{x \cdot 0}{x^2 + 0} = 0,$$

$$\underset{y \to 0}{L} \ \underset{x \to 0}{L} \frac{xy}{x^2 + y^2} = \underset{y \to 0}{L} \frac{0 \cdot x}{0 + y^2} = 0,$$

and hence the order of limits may be interchanged.

A necessary and sufficient condition for the interchange of the order of iterated limits may be stated as follows.

THEOREM VI. *A necessary and sufficient condition that*

$$\underset{x\to x_0}{L}\ \underset{y\to y_0}{L} f(x, y) = \underset{y\to y_0}{L}\ \underset{x\to x_0}{L} f(x, y)$$

*is:*

(*a*) $$\underset{x\to x_0}{L} f(x, y) = \phi(y),\ y \neq y_0,$$

$$\underset{y\to y_0}{L} f(x, y) = \psi(x),\ x \neq x_0;$$

(*b*) $$\underset{y\to y_0}{L} \phi(y) = A;$$

(*c*) *There exists a set of values* $y = \{y_n\}$ *dense at* $y = y_0$ *such that for each* $y_n$, $n > m$, *there exists on the line* $y = y_n$ *an interval* $(x_0 - \delta_{y_n}, x_0 + \delta_{y_n})$ *which may vary with* $y_n$, *and for every* $x$ *within this interval, we have*

$$|f(x, y_n) - \psi(x)| < \epsilon.$$

We shall first show that the condition is necessary; that is, whenever the interchange of limits is permissible the condition set forth in the theorem follows. From the assumed relation,

$$\underset{x\to x_0}{L}\ \underset{y\to y_0}{L} f(x, y) = \underset{y\to y_0}{L}\ \underset{x\to x_0}{L} f(x, y) = A,$$

the conditions (*a*) and (*b*) follow at once. We must show that condition (*c*) is also satisfied. We have

$$\underset{x\to x_0}{L}\ \underset{y\to y_0}{L} f(x, y) = \underset{x\to x_0}{L} \psi(x) = A,$$

which gives, when stated in the form of an inequality,

$$|\psi(x) - A| < \frac{\epsilon}{3},\ |x - x_0| < \delta_1. \tag{7}$$

We have also $$\underset{y\to y_0}{L}\ \underset{x\to x_0}{L} f(x, y) = \underset{y\to y_0}{L} \phi(y) = A,$$

and hence we may write for any set $\{y_n\}$ dense at $y_0$

$$|A - \phi(y_n)| < \frac{\epsilon}{3},\ |y_n - y_0| < \delta_2. \tag{8}$$

Moreover, we have the limit

$$\underset{x\to x_0}{L} f(x, y_n) = \phi(y_n)$$

and hence the inequality

$$|\phi(y_n) - f(x, y_n)| < \frac{\epsilon}{3},\ |x - x_0| < \delta_3. \tag{9}$$

Combining the inequalities (7), (8), and (9), we obtain

$$|\psi(x) - f(x, y_n)| < \epsilon, \ |x - x_0| < \delta_{y_n},$$

where $\delta_{y_n}$ is the smaller of the two numbers $\delta_1$, $\delta_3$. This relation is that set forth in condition (*c*) of the theorem.

The conditions of the theorem are also sufficient. The first of the conditions (*a*) holds for all values of $x$ in every deleted neighborhood of $x_0$, say in $(x_0 - \delta'_{y_n}, x_0 + \delta'_{y_n})$. Hence, for the values of $x$ in this neighborhood and for each value of $y_n$ in $\{y_n\}$ we have

$$|f(x, y_n) - \phi(y_n)| < \frac{\epsilon}{3}. \tag{10}$$

Since the limit (*b*) exists, then for values of $y_n$ sufficiently close to $y_0$ we have

$$|\phi(y_n) - A| < \frac{\epsilon}{3}. \tag{11}$$

For each $y_n$ we have from (*c*) for all values of $x$ in an interval $(x_0 - \delta_{y_n}, x_0 + \delta_{y_n})$ the relation

$$|\psi(x) - f(x, y_n)| < \frac{\epsilon}{3}. \tag{12}$$

Hence, having chosen a $y_n$ sufficiently close to $y_0$, we have upon combining (10), (11) and (12) the relation

$$|\psi(x) - A| < \epsilon, \tag{13}$$

which holds for all values of $x$ in the deleted neighborhood of $x_0$ defined by the smaller of the intervals $(x_0 - \delta_{y_n}, x_0 + \delta_{y_n})$, $(x_0 - \delta'_{y_n}, x_0 + \delta'_{y_n})$. From this relation, it follows that

$$\underset{x \to x_0}{L} \psi(x) = A,$$

which is equivalent to saying that the order of the two iterated limits given in the theorem may be interchanged.

If we now examine more closely the character of the function given in Ex. 2, we shall see why the iterated limits can be interchanged in that case. Conditions (*a*) and (*b*) of the foregoing theorem are satisfied; for, we have

$$\underset{x \to 0}{L} \frac{xy}{x^2 + y^2} = 0 = \phi(y), y \neq 0,$$

$$\underset{y \to 0}{L} \frac{xy}{x^2 + y^2} = 0 = \psi(x), x \neq 0,$$

$$\underset{y \to 0}{L} \phi(y) = 0.$$

The third condition of the theorem is also satisfied. On each line $y = y_n$, where $\{y_n\}$ is any set of values of $y$ dense at $y = 0$ there is an interval $(0, \delta_{y_n})$, which varies with $y_n$, such that for each value of $x$ in the interval, we have

$$| f(x, y_n) - \phi(y_n) | < \epsilon.$$

As the condition (b) is satisfied, we have

$$| \phi(y_n) - 0 | < \epsilon, \; | y_n | < \delta'.$$

From the fact that $f(x, y) = 0$ for all values of $x$ on the line $y = 0$, we have

$$| 0 - \psi(x) | < \epsilon.$$

By combining these three inequalities we have the desired result

$$| f(x, y_n) - \psi(x) | < 3\epsilon,$$

for all values of $x$ within the interval $(0, \delta_{y_n})$. Hence, the conditions of the theorem are all satisfied, and the order of the limits may be interchanged as we have seen.

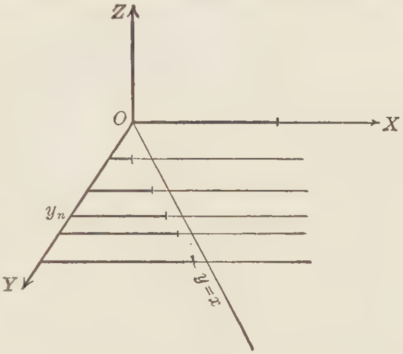


FIG. 15.

In certain theoretical discussions it is useful to have the following form of the necessary and sufficient condition for the existence of a limit of a function.

THEOREM VII. *A necessary and sufficient condition that the limit* $\underset{x \to x_0}{L} f(x)$ *exists is that for an arbitrarily small positive number* $\epsilon$ *there exists a positive number* $\delta$ *such that for any two points* $x, \bar{x}$, *distinct from* $x_0$, *in the interval* $(x_0 - \delta, x_0 + \delta)$ *we have*

$$| f(x) - f(\bar{x}) | < \epsilon. \tag{a}$$

That the condition is necessary follows at once. For, if the limit $\underset{x \to x_0}{L} f(x)$ exists, we may write

$$| f(x) - A | < \frac{\epsilon}{2}, \; | x - x_0 | < \delta,$$

$$| A - f(\bar{x}) | < \frac{\epsilon}{2}, \; | \bar{x} - x_0 | < \delta.$$

Combining these inequalities, we have

$$| f(x) - f(\bar{x}) | < \epsilon, \; | x - x_0 | < \delta, \; | \bar{x} - x_0 | < \delta.$$

To show that the given condition is also sufficient, we assume that the inequality (a) holds for all values of $x$ and $\bar{x}$, other than $x_0$ itself, in the interval $(x_0 - \delta, x_0 + \delta)$. On the basis of this assumption, we must show that there exists a definite point $A$ on the line $x = x_0$

toward which the values of $f(x)$ tend as $x$ is made to approach $x_0$; that is, we must have

$$|f(x) - A| < \epsilon, \ |x - x_0| < \delta.$$

Let $\{\eta_k\}$ be a sequence of decreasing, positive numbers such that

$$\underset{k \to \infty}{L} \eta_k = 0.$$

Then, from ($a$) it follows that if we assign the value $\eta_1$ to the arbitrary number $\epsilon$, there must exist a positive number $\delta$, such that we have

$$|f(x) - f(\bar{x})| < \eta_1,$$

where $x_0 - \delta_1 \leqq x \neq x_0 \leqq x_0 + \delta_1, \ x_0 - \delta_1 \leqq \bar{x} \neq x_0 \leqq x_0 + \delta_1.$

Then, if $a_1$ is any point other than $x_0$ in the interval $(x_0 - \delta_1, x_0 + \delta_1)$, it follows that the values of $f(x)$ for any $x \neq x_0$ in this interval must lie between $f(a_1) - \eta_1$ and $f(a_1) + \eta_1$. The lines

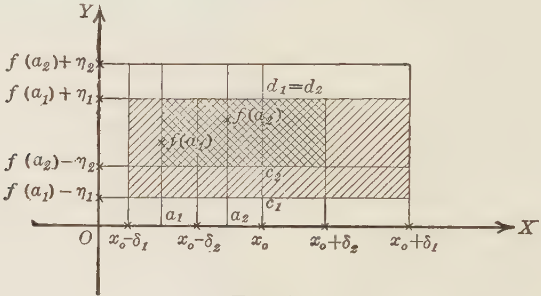


FIG. 16.

$$y = f(a_1) - \eta_1, \ y = f(a_1) + \eta_1$$

cut from the line $x = x_0$ an interval $(c_1, d_1)$ whose length is $2\eta_1$.

Again, if we assign to $\epsilon$ the value $\eta_2 < \eta_1$, we have in the same way an interval $(x_0 - \delta_2, x_0 + \delta_2)$ such that for values of $x \neq x_0$ in this interval the corresponding values of $f(x)$ must lie between $f(a_2) - \eta_2$, $f(a_2) + \eta_2$, $f(a_2)$ being the functional value at some point $a_2 \neq x_0$ in $(x_0 - \delta_2, x_0 + \delta_2)$. The lines

$$y = f(a_2) - \eta_2, \ y = f(a_2) + \eta_2$$

determine on $x = x_0$ a second interval $(c_2, d_2)$ of length $2\eta_2$. If it happens that either extremity of this interval falls outside the interval $(c_1, d_1)$ we select as the interval $(c_2, d_2)$ that portion common to the two intervals, as indicated in Fig. 16, in which case the length of the interval is less than $2\eta_2$. In a similar manner by assigning to $\epsilon$ the value $\eta_3 < \eta_2$ we are led to a third interval $(c_3, d_3)$ on $x = x_0$ of length not greater than $2\eta_3$ and lying within the interval $(c_2, d_2)$. Continuing this process we obtain on $x = x_0$ a sequence of intervals

$$(c_1, d_1), (c_2, d_2), (c_3, d_3), \cdot \cdot \cdot , (c_k, d_k), \cdot \cdot \cdot \tag{14}$$

each lying within the preceding. The limit of the lengths of these intervals is zero, since $(c_k, d_k)$ can not exceed $2\eta_k$, which has the limit zero. The sequence (14) then defines uniquely a point $A$ on $x = x_0$. This point $A$ lies within all of the intervals $(c_k, d_k)$ however large $k$ may be chosen.

Corresponding to each interval $(c_k, d_k)$ there is on the $X$-axis a definite interval $(x_0 - \delta_k, x_0 + \delta_k)$ of length greater than zero such that for each point $x$ within it the value of $f(x)$ differs numerically from the value of the ordinate at any other point of $(x_0 - \delta_k, x_0 + \delta_k)$ by less than $\eta_k$, provided the upper extremity of that ordinate lies between the lines

$$y = f(a_k) - \eta_k \text{ and } y = f(a_k) + \eta_k.$$

As the point $A$ on the line $x = x_0$ satisfies this condition, we may write

$$| f(x) - A | < \eta_k, \;\; x_0 - \delta_k \leqq x \neq x_0 \leqq x_0 + \delta_k,$$

where $\eta_k$ may be made arbitrarily small by the proper selection of $k$, and $\delta_k$ is different from zero. Replacing the arbitrarily small number $\eta_k$ by $\epsilon$ and $\delta_k$ by $\delta$, we can express the foregoing result in the required form, thus establishing the theorem.

The corresponding theorem for simultaneous limits involving two independent variables may be stated as follows. The proof is similar to that of the foregoing theorem.

THEOREM VIII. *A necessary and sufficient condition that the simultaneous limit* $\underset{\substack{x \to x_0 \\ y \to y_0}}{\mathrm{L}} f(x, y)$ *exists is that for an arbitrarily small positive number* $\epsilon$ *there exists a positive number* $\delta$ *and a region*

$$(x_0 - \delta \leqq x \leqq x_0 + \delta, \;\; y_0 - \delta \leqq y \leqq y_0 + \delta)$$

*such that for any two points* $(x, y)$, $(\bar{x}, \bar{y})$ *in this region exclusive of* $(x_0, y_0)$ *itself we have*

$$| f(x, y) - f(\bar{x}, \bar{y}) | < \epsilon.$$

**25. Continuity and discontinuity of a function.** Closely associated with the notion of a limit is that of the continuity of a function. We say that $f(x)$ is **continuous**[1] at the point $x = x_0$ if we have

$$\underset{x \to x_0}{L} f(x) = f(x_0),$$

or what is the same thing, $f(x)$ is continuous at $x_0$ if for an arbitrarily small positive number $\epsilon$ there exists a positive number $\delta$ such that

$$|f(x) - f(x_0)| < \epsilon, \; |x - x_0| < \delta. \tag{1}$$

This definition involves the condition that the limit shall exist and that its value shall be the value of the function at the point in question. As we have seen, the mere existence of a limit does not require that the function be defined at the limiting point, since it depends only on the values of the function in the deleted neighborhood of that point. For the continuity of a function, however, it must be defined at the limiting point.

The function $\sin \frac{1}{x}$ is not defined for $x = 0$, since for this value it does not have a definite value. If we assign to it some value at this point, say the value zero, we may then consider its continuity at that point. This function is discontinuous at the origin; for, as we have seen, the limit $\underset{x \to 0}{L} \sin \frac{1}{x}$ does not exist. On the other hand, the function $x \sin \frac{1}{x}$ does have a limit, namely zero, as $x$ approaches zero. If we assign the value zero to the function at the origin, it is continuous. Discontinuity may also occur if the limiting values are different on the two sides of the critical point as the following illustration shows.

**Ex. 1.** Consider the continuity for $x = 0$ of $f(x) = \dfrac{1}{1 - e^{\frac{1}{x}}}$.

The right-hand limit gives $\underset{x \to 0+}{L} \dfrac{1}{1 - e^{\frac{1}{x}}} = 0,$

while the left-hand limit gives

$$\underset{x \to 0-}{L} \frac{1}{1 - e^{\frac{1}{x}}} = \underset{x \to 0+}{L} \frac{1}{1 - \dfrac{1}{e^{\frac{1}{x}}}} = 1.$$

Since these two limiting values are different, the limit as $x$ approaches zero can not be said to exist. Hence, the given function is discontinuous at the origin.

---

[1] The first precise definition of continuity was given by Cauchy in his *Analyse algebraic* (1821).

**Ex. 2.** Consider the continuity of $f(x) = \underset{t \to 0}{L}\, x^{2t}$, where $t \neq \frac{1}{4n}$, $n$ being an integer.

This function is fully defined, as it is 0 for $x = 0$ and $+ 1$ for all other values of $x$. Here, the limit $\underset{x \to 0}{L}\, f(x)$ exists and is equal to $+ 1$, but as that is different from the value of $f(x)$ at $x = 0$, the given function is discontinuous.

If at a point of discontinuity $x_0$ the right-hand and the left-hand limits of the function both exist but are distinct; that is, if

$$f(x_0 + 0) \equiv \underset{x \to x_0+}{L} f(x) \neq f(x_0 - 0) \equiv \underset{x \to x_0-}{L} f(x),$$

then $f(x)$ is said to have an **ordinary discontinuity,** or a discontinuity of the first kind.

If we have $$f(x_0 + 0) = f(x_0 - 0) \neq f(x_0),$$

then the discontinuity is often spoken of as a **removable discontinuity;** for, we can always make $x_0$ a point of continuity by assigning to $f(x)$ at that point the common value of the right-hand and the left-hand limits.

If either the right-hand limit or the left-hand limit or both fail to exist at $x_0$, then the function is said to have at $x_0$ a **discontinuity of the second kind,** no matter what value be assigned to the function at this point. Thus the function

$$f(x) = \sin \frac{1}{x}, \quad x \neq 0,$$
$$= 0, \qquad x = 0$$

has such a discontinuity at the origin.

If the limit of the values of the function on the right of the limiting point $x = x_0$ is the same as the value of the function at the point, the function is said to be **continuous on the right** of $x_0$. In a similar manner a function may be **continuous on the left** of $x_0$. In order to be continuous at the point $x_0$ the function must be continuous both on the right and on the left of the point. Of course a function may be continuous on one side and have a discontinuity on the other due to the non-existence of the limit of the function or to the fact that the function has a value at the point different from the value of the limit. Likewise it may have a discontinuity on one side due to the non-existence of the limit, and on the other side a discontinuity due to the difference of the limit and the value of the function at the point. In other words, a function may be continuous on the one side and have a discontinuity of either the first kind or second kind on the other side, or it may have a discontinuity of the first kind on the one side and one of the second kind on the other side.

A function is **continuous in an interval** if it is continuous at all points of the interval. If $f(x)$ is continuous in a closed interval $(a, b)$, then the least upper bound of $f(x)$ in $(a, b)$ is identical with its upper limit in this closed interval. So likewise the greatest lower bound and the lower limit are identical. On the other hand, if $f(x)$ is known to be continuous merely in the open interval $(a, b)$, then one can not make the foregoing statement concerning the equality of the upper limit and the least upper bound in the closed interval. In fact, there may, in this case, be neither a finite upper limit nor a finite upper bound, or both may be finite and different. A similar statement may be made concerning the lower limit and the greatest lower bound.

We shall have occasion to measure the amount of discontinuity which a bounded function may have at a given point $x_0$. Suppose we have an interval $(a, b)$ containing the point $x_0$. The function $f(x)$, for $a \leqq x \leqq b$, has a least upper bound which we shall denote by $M(f, ab)$. Similarly, the greatest lower bound of $f(x)$ for this same interval exists and may be denoted by $m(f, ab)$. The **oscillation** $\omega(f, ab)$ of $f(x)$ in the closed interval $(ab)$ may then be defined by the equation

$$\omega(f, ab) = M(f, ab) - m(f, ab). \tag{2}$$

Consider now a neighborhood of the point $x_0$, say $(x_0 - \delta, x_0 + \delta)$, lying within the interval $(a, b)$. As $\delta$ is allowed to decrease, the value of $M(f, \delta)$ never increases, and that of $m(f, \delta)$ never decreases. Moreover, we have for all values of $x$ in this interval

$$M(f, \delta) \geqq f(x_0), \quad m(f, \delta) \leqq f(x_0);$$

that is, both $M(f, \delta)$ and $m(f, \delta)$ are not only monotone, but the first has a finite lower bound and the second a finite upper bound. Both have a limit as $\delta$ approaches zero and we shall denote these limits by $M(f, x_0)$, $m(f, x_0)$, respectively. The two values $M(f, x_0)$, $m(f, x_0)$ are often spoken of as the **maximum** and the **minimum** values, respectively, of $f(x)$ at the point $x_0$. The maximum value $M(f, x_0)$ is the largest of the three numbers, $f(x_0)$, the upper right-hand limit $\overline{f(x_0 + 0)}$, and the upper left-hand limit $\overline{f(x_0 - 0)}$. Similarly, the minimum value at $x_0$, namely $m(f, x_0)$ is the smallest of the three numbers, $f(x_0)$, $\underline{f(x_0 + 0)}$, $\underline{f(x_0 - 0)}$.

The relation

$$\omega(f, x_0) \equiv M(f, x_0) - m(f, x_0) \tag{3}$$

is important in telling us of the behavior of $f(x)$ as to continuity at $x_0$. If the interval $(x_0 - \delta, x_0 + \delta)$ is regarded as the deleted neighborhood of $x_0$, then $\omega(f, x_0)$ is called the **oscillation of $f(x)$ at the point $x_0$.** In this case the value of $\omega(f, x_0)$ is not affected by any value which may be assigned to $f(x)$ at $x_0$ or indeed whether the function is even defined at that point. If $f(x)$ is defined at $x_0$ and its value at this point is taken into account in determining $M(f, x_0)$, $m(f, x_0)$, we shall then speak of $\omega(f, x_0)$ as the **saltus of $f(x)$ at $x_0$.**[1]

It follows at once that if the saltus at a point is zero the function is continuous at that point, and conversely if it is continuous the saltus is necessarily zero. Neither the oscillation nor the saltus is ever negative. For example, the oscillation of $\sin \frac{1}{x}$ at $x = 0$ is 2, and for the function $\frac{1}{1 - e^{\frac{1}{x}}}$ it is 1 at the origin. For each point $x_0$ of an interval there are then defined three numbers $M(f, x_0)$, $m(f, x_0)$, $\omega(f, x_0)$, satisfying the relation (3).

We have thus far spoken of the continuity of a function at a point with respect to the entire continuum. We may also speak of the continuity of $f(x)$ with respect to a given point set $E$ for which it is defined. Thus, $f(x)$ is continuous at the point $x_0$ with respect to the set of points $E$, if the definition of continuity is satisfied as $x$ ranges over this set rather than over the continuum. In order that it be continuous at each point of $E$, each such point must be itself a limiting point of $E$, that is $E$ must be dense in itself; for, it is only at such points that the notion of continuity can have any significance, since at an isolated point the limit can not be said to exist. The set $E$ need not be closed; for, at those limiting points of $E$ which are not contained in $E$ the given function is not defined. Likewise the set $E$ need

---

[1] The distinction between the oscillation at a point and the saltus may also be made clear by use of the upper and lower right-hand and left-hand limits

$$\overline{L}_{x \to x_0+} f(x) = \overline{f(x_0 + 0)}, \qquad \overline{L}_{x \to x_0-} f(x) = \overline{f(x_0 - 0)},$$
$$\underline{L}_{x \to x_0+} f(x) = \underline{f(x_0 + 0)}, \qquad \underline{L}_{x \to x_0-} f(x) = \underline{f(x_0 - 0)}.$$

Thus the oscillation at $x_0$ then becomes the excess of the largest over the smallest of the four numbers $\overline{f(x_0 + 0)}$, $\overline{f(x_0 - 0)}$, $\underline{f(x_0 + 0)}$, $\underline{f(x_0 - 0)}$, while the saltus is the excess of the largest over the smallest of the five numbers $\overline{f(x_0 + 0)}$, $\overline{f(x_0 - 0)}$, $\underline{f(x_0 + 0)}$, $\underline{f(x_0 - 0)}$, $f(x_0)$. [Cf. Hobson *Functions of a Real Variable*, 2d Ed. p. 284.] The *Encyclop. des Sci. Math.* uses the term *vibration* for oscillation at a point, [II$_2$, p. 62]. Baire uses "oscillation at a point" in the sense of saltus as defined above [see *Leçons sur les fonctions discontinues*, p. 71].

not be everywhere dense in the given interval; for, we may have continuity relative to any perfect set and, as we have seen (Art. 17), such a set need not be everywhere dense in any portion of a given interval.

A function $f(x)$ is said to be **absolutely continuous**[1] in a given interval $(a, b)$, if for every finite or enumerably infinite set of non-overlapping intervals in $(a, b)$ the sum of the oscillations of $f(x)$ in these intervals tends to zero with the sum of the lengths of the intervals. From this definition it follows that any function which is absolutely continuous is also continuous in the ordinary sense. The converse of this statement, however, is not true. For example, the function

$$\begin{aligned} f(x) &= x \sin \frac{1}{x}, \quad \text{for } x \neq 0, \\ &= 0, \qquad\quad \text{for } x = 0 \end{aligned}$$

is continuous in the interval $(0, 1)$, but it is not absolutely continuous. In each of the subintervals $\left(\frac{1}{k\pi}, \frac{1}{(k+2)\pi}\right)$, $k = 1, 2, 3, \cdots$, the oscillation of $f(x)$ exceeds $\frac{2}{(k+2)\pi}$. The sum of the oscillations in the enumerably infinite set of subintervals to be found in the neighborhood $\left(0, \frac{1}{n\pi}\right)$ of the origin, is greater than

$$\frac{2}{(n+2)\pi} + \frac{2}{(n+3)\pi} + \cdots = \frac{2}{\pi}\left\{\frac{1}{n+2} + \frac{1}{n+3} + \cdots\right\}.$$

But, however large $n$ may be chosen, and hence however small the neighborhood may be, the sum of the oscillations exceeds any finite bound, since the infinite series in the brackets is divergent for every value of $n$. Consequently, the given function can not be absolutely continuous.

The notion of absolute continuity can be extended also to functions defined for a set of points. Thus $f(x)$ is absolutely continuous on the set of points $E$, provided that for every subset $E_1$ of $E$ the sum of the oscillations of $f(x)$ tends to the limit zero with the measure of $E_1$.

Suppose we have given a function $f(x)$ which is continuous throughout the interval $(a, b)$. For a given, arbitrarily small positive number $\epsilon$, there is then determined a positive number $\delta$, which may vary with $x_0$, such that the conditions set forth by the inequality (1) are satisfied for each constant value $x_0$ in $(a, b)$. For a fixed value of $\epsilon$ this $\delta$ is, as already pointed out, a definitely defined function of $x_0$. The aggregate of values of $\delta(x_0)$ as $x_0$ takes all values in the interval $(a, b)$ has a greatest lower bound. If, for a fixed but arbitrarily small value of $\epsilon$, the lower bound of $\delta(x_0)$ is different from zero, then $f(x)$ is said to be **uniformly continuous** in the interval $(a, b)$; for, this lower bound $\delta'$ may be associated with the chosen $\epsilon$ uniformly for all values of $x_0$ in $(a, b)$. In other words, there exists in this case a $\delta$ independent of $x_0$ which together with $\epsilon$ satisfies the conditions of the inequality (1) defining continuity for each choice of $x_0$ in the given interval.

---

[1] Such functions were first defined and their properties discussed by Vitali. See *Atti della R. Accad. delle Scienze di Torino* (1905), Vol. 40, p. 1021.

As we have seen, a given function may be discontinuous at a single point. As the following illustrations show, it may also be discontinuous at a set of points.

**Ex. 3.** Given the function

$$f(x) = \underset{t\to\infty}{L} \frac{(1+\sin \pi x)^t - 1}{(1+\sin \pi x)^t + 1}.$$

Show that $f(x)$ is discontinuous at the points

$$x = 0, 1, 2, 3, \cdot \cdot \cdot, n, \cdot \cdot \cdot.$$

For $x = n$, we have $\quad f(x) = \underset{t\to\infty}{L} \dfrac{(1+0)^t - 1}{(1+0)^t + 1} = 0.$

For $x = x_0$, where $2n < x_0 < 2n + 1$, we have

$$0 < \sin \pi x_0 \leqq 1,$$

and hence $\quad f(x) = \underset{t\to\infty}{L} \dfrac{(1+\sin \pi x_0)^t - 1}{(1+\sin \pi x_0)^t + 1} = \underset{t\to\infty}{L} \dfrac{1 - \dfrac{1}{(1+\sin \pi x_0)^t}}{1 + \dfrac{1}{(1+\sin \pi x_0)^t}} = 1.$

On the other hand, for $x = x_0$, where $2n + 1 < x_0 < 2n + 2$, we have

$$-1 \leqq \sin \pi x_0 < 0,$$

and hence $\quad f(x) = \underset{t\to\infty}{L} \dfrac{(1+\sin \pi x_0)^t - 1}{(1+\sin \pi x_0)^t + 1} = -1.$

The function therefore has a discontinuity of the first kind at $x = 0, 1, 2, \ldots,$ and the graph is given in Fig. 17.

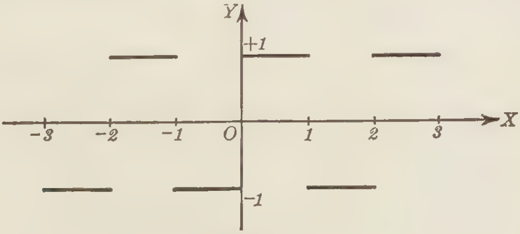


FIG. 17.

**Ex. 4.** Given the function $f(x) = \underset{n\to\infty}{L}\ \varphi_n(x)$, where

$$\varphi_n(x) = \underset{t\to 0}{L} \frac{\sin^2 n!\,\pi x}{\sin^2 n!\,\pi x + t^2}.$$

Show that $f(x)$ is totally discontinuous.

Let us first consider $f(x)$ for rational values of $x$; that is, let $x = \frac{p}{q}$, where $p$ and $q$ are integers prime to each other. By taking $n$ sufficiently large $n!\,\pi x$ is an integral multiple of $\pi$ for any given value of $x$. We need at most to put $n = q$. For any such value of $x$, we have

$$\sin n!\,\pi x = 0,$$

and hence
$$f(x) = \underset{n\to\infty}{L}\ \underset{t\to 0}{L} \frac{\sin^2 n!\,\pi x}{\sin^2 n!\,\pi x + t^2} = \underset{n\to\infty}{L}\ \underset{t\to 0}{L} \frac{0}{0 + t^2} = 0.$$

If $x$ is irrational, we have $x \neq \frac{p}{q}$ and hence for all values of $n$ however large we have
$$\sin^2 n!\,\pi x > 0.$$

$$f(x) = \underset{n\to\infty}{L}\ \underset{t\to 0}{L} \frac{\sin^2 n!\,\pi x}{\sin^2 n!\,\pi x + t^2} = \underset{n\to\infty}{L}\ \underset{t\to 0}{L} = \frac{1}{1 + \dfrac{t^2}{\sin^2 n!\,\pi x}} = \frac{1}{1+0} = 1.$$

Since $f(x)$ is 0 for all rational values of $x$ and is 1 for all irrational values, it is totally discontinuous; that is, it is discontinuous for every value of $x$.

In between the class of functions which, like the foregoing, are totally discontinuous and the class of continuous functions there is a class called **point-wise discontinuous functions**.[1] A discontinuous function is said to be point-wise discontinuous in a given interval if the points of continuity are everywhere dense but do not form a closed set. Such a function has then points of continuity in every interval however small, but in every such interval there may be also points of discontinuity. The class of point-wise discontinuous functions forms a very important class and their properties will be frequently used in the discussions to follow.

Thus far we have discussed the continuity of functions of a single variable. The consideration of the continuity of functions of several independent variables involves simultaneous limits. Thus, we say that $f(x, y)$ is continuous at the point $(x_0, y_0)$ if

$$\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x, y) = f(x_0, y_0).$$

---

[1] First discussed by Hankel, who was the first to classify functions as continuous, point-wise discontinuous, and totally discontinuous and to discuss the characteristic properties of discontinuous functions. See, *Untersuchungen über die unendlich oft oscillierenden und unstetigen Funktionen*, Tübingen (1870); also *Math. Ann.*, Vol. 20 (1882), p. 89.

By continuity here is meant, of course, continuity with respect to the two variables taken together. Stated in the form of an inequality, $f(x, y)$ is continuous if for an arbitrarily chosen $\epsilon$, there is determined a positive number $\delta$ such that

$$|f(x, y) - f(x_0, y_0)| < \epsilon, |x - x_0| < \delta, |y - y_0| < \delta.$$

Instead of being continuous with respect to $x$ and $y$ together, $f(x, y)$ may be continuous in $x$ and in $y$ separately; that is, the functions may be continuous in $x$ for constant values of $y$ or continuous in $y$ for constant values of $x$. Continuity of this latter kind does not involve continuity with respect to the two variables together. However, if a function is continuous in the two variables together it is of course continuous with respect to each variable separately.

**Ex. 5.** Consider the continuity at the point (0, 0) of $f(x, y) = \dfrac{xy^3}{x^2 + y^6}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$.

The simultaneous limit $\underset{\substack{x \to 0 \\ y \to 0}}{L} f(x, y)$ does not exist; for, if we allow the variables to approach the limiting point by first permitting $x$ to approach zero while $y$ remains constant and then permit $y$ to approach zero we get the limiting value zero. On the other hand, if we permit the variables to approach zero along the path $y^3 = x$, we have

$$\underset{y \to 0}{L} \frac{y^6}{y^6 + y^6} = \frac{1}{2}.$$

Consequently, $f(x, y)$ is not continuous at (0, 0) in $(x, y)$ together. It is, however, continuous in $x$ alone and in $y$ alone at the origin. For, putting either variable equal to zero and then allowing the other to approach zero we have the limiting value zero, which is the value of $f(0, 0)$.

In order that $f(x, y)$ shall be continuous in both variables together, it must have the same limiting value by all possible approaches to the critical point. A necessary and sufficient condition for continuity therefore involves the condition that the function is not only continuous in each direction but the continuity is uniform for all directions. For, if we put

$$x = x_0 + \rho \cos \theta, y = y_0 + \rho \sin \theta,$$

we have from the definition of continuity

$$|f(x_0 + \rho \cos \theta, y_0 + \rho \sin \theta) - f(x_0, y_0)| < \epsilon,$$

which must hold for all values of $\rho$ less than some number $\rho_0$ which is independent of $\theta$. This is equivalent to saying that the transformed function must be uniformly continuous in $\rho$ for all values of $|\theta| \leqq 2\pi$. The following example will illustrate this condition.

**Ex. 6.** Consider the continuity at (0, 0) of

$$f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}, \; x \neq 0, \; y \neq 0, \text{ where } f(0, 0) = 0.$$

Putting $x = \rho \cos \theta$, $y = \rho \sin \theta$, we obtain

$$f(\rho \cos \theta, \rho \sin \theta) = \frac{\rho^2 \cos \theta \sin \theta}{\rho\sqrt{\sin^2 \theta + \cos^2 \theta}} = \frac{1}{2} \rho \sin 2\theta,$$

which is arbitrarily small if we take

$$\rho < \frac{2\epsilon}{\sin 2\theta} = 2\epsilon \cdot \csc 2\theta.$$

Since $\csc 2\theta$ is never less than one, it follows that if we put $\rho_0 = 2\epsilon$, then however small $\epsilon$ may be chosen $f(\rho \cos \theta, \rho \sin \theta)$ is always less than $\epsilon$ for all values of $\theta$ and for $\rho < \rho_0$; that is, at all points within a circle about the origin and with a radius $\rho_0 = 2\epsilon$. The transformed function is, therefore, uniformly continuous in $\rho$ for all values of $\theta$. Hence, the foregoing criterion is satisfied and $f(x, y)$ is continuous in $(x, y)$ together at the origin.

It must not be inferred that because

$$\underset{\rho\to 0}{L} f(x_0 + \rho \cos \theta, y_0 + \rho \sin \theta) = f(x_0, y_0)$$

for each value of $\theta$, the function is necessarily continuous. For example, it has been shown (Ex. 5) that the simultaneous limit of

$$f(x, y) = \frac{xy^3}{x^2 + y^6}, \; x \neq 0, \; y \neq 0, \text{ where } f(0, 0) = 0,$$

as $x$ and $y$ approach zero does not exist. This function is, therefore, discontinuous at the origin. However, we have after introducing polar coördinates

$$\underset{\rho\to 0}{L} \frac{\rho^4 \cos \theta \sin^3 \theta}{\rho^2 \cos^2\theta + \rho^6 \sin^6\theta} = \underset{\rho\to 0}{L} \rho^2 \frac{\cos \theta \sin^3 \theta}{\cos^2 \theta + \rho^4 \sin^6 \theta} = 0 = f(0, 0)$$

for each constant value of $\theta$.

The terms oscillation, uniform continuity, point-wise discontinuous already discussed for functions of a single variable can be at once extended to functions of two or more independent variables. All that is necessary is to replace the term interval by region in the previous definitions, it being understood also that continuity here has reference to the several variables taken together rather than to that of each separately. The measure of discontinuity or saltus of a function at a point is, as in the case of a function of a single variable, the difference between the maximum and the minimum values of the function at the point in question. It is always equal to or greater than the difference between any two limiting values at the point.

**Ex. 7.** Show that the saltus of the function $f(x, y) = \frac{xy}{x^2 + y^3}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$, is infinitely large at the origin.

We have from the iterated limits

$$\underset{x \to 0}{L} \ \underset{y \to 0}{L} \frac{xy}{x^2 + y^3} = \underset{x \to 0}{L}\ 0 = 0,$$

while putting $x = y^{\frac{3}{2}}$ and then passing to the limit, we have

$$\underset{y \to 0}{L} \frac{y^{\frac{5}{2}}}{2y^3} = \underset{y \to 0}{L} \frac{1}{2y^{\frac{1}{2}}} = \infty.$$

**26. Properties of continuous functions.** Having defined a continuous function in terms of a limit, the laws of operation with limits apply to such functions. Thus the sum, difference, and product of a finite number of continuous functions is a continuous function. The quotient of one continuous function by another is continuous at any point at which the denominator is not zero. It follows also that any rational function is continuous at all points where its denominator does not vanish. Functions which are continuous in a closed interval satisfy the following theorem.

THEOREM I. *If a function $f(x)$ is continuous in a closed interval, it is uniformly continuous in that interval.*

From the continuity of $f(x)$ we have

$$| f(x) - f(x_0) | < \epsilon, \ | x - x_0 | < \delta.$$

The value of $\delta$ depends upon $\epsilon$ and also upon the choice of $x_0$. If $(a, b)$ is the given interval, then for each value of $\epsilon$ there is associated with every point $x$ in $(a, b)$ a definite value $\delta(x)$. The aggregate of all these values has a greatest lower bound $\delta'$, which we wish to show is different from zero under the conditions set forth in the theorem. Since the interval $(a, b)$ is closed it must contain at least one point such that in every neighborhood, however small, of this point the greatest lower bound of $\delta(x)$ is $\delta'$. Let $x_1$ be such a point. By hypothesis $f(x)$ is continuous at $x_1$ and hence we may write

$$| f(x) - f(x_1) | < \frac{\epsilon}{2},$$

which is satisfied by each value of $x$ in an interval $(x_1 - \delta_1, x_1 + \delta_1)$. Then, in the interval $\left(x_1 - \frac{\delta_1}{2}, x_1 + \frac{\delta_1}{2}\right)$ the oscillation of $f(x)$ can, of course, not exceed $\epsilon$. Moreover, for a given $\epsilon$, there is associated with each point $x$ of this interval a value of $\delta(x)$ which is at least equal to $\frac{\delta_1}{2}$. The greatest lower bound of $\delta(x)$ in this interval is therefore equal to or greater than $\frac{\delta_1}{2}$, which is, as we have seen, greater than zero. Since the greatest lower bound of $\delta(x)$ for the interval $(a, b)$, namely $\delta'$, is the greatest lower bound of $\delta(x)$ in every arbitrarily small neighborhood of $x_1$, it follows that we have

$$\delta' \geqq \frac{\delta_1}{2} > 0,$$

which establishes the theorem.

It is essential to the validity of this theorem that the interval $(a, b)$ of continuity be closed. If it is not, the greatest lower bound of the $\delta(x)$ may be zero. Such is the case for the function

$$f(x) = \frac{1}{x}, \quad 0 < x \leqq 1.$$

This function is continuous in the given interval, and hence for any arbitrarily chosen $\epsilon$ there exists for each $x$ a $\delta(x) > 0$ equal to the smaller of the two numbers $\delta_1(x)$, $\delta_2(x)$, as shown in Fig. 18. The function $\delta(x)$ has the greatest lower bound zero as $x$ takes all values in the neighborhood of the origin, as may be seen from the graph of the function.

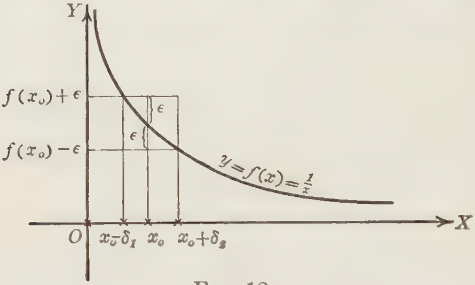


FIG. 18.

THEOREM II. *If $f(x)$ is continuous in a closed, finite interval $(a, b)$, then this interval can be divided into a finite number of subintervals in each of which the oscillation of $f(x)$ is less than an arbitrarily chosen number $\epsilon$.*

From Theorem I it follows that however small $\epsilon$ may be selected there exists a positive number $\delta$ independent of $x$ such that the oscillation of $f(x)$ in every interval $(x - \delta, x + \delta)$ lying in $(a, b)$ is less than $\epsilon$. Since $\delta$ is greater than zero and any point $x$ of $(a, b)$ can be taken as the center, it follows that a finite number of such subintervals is sufficient to cover the given interval $(a, b)$.

This theorem also follows as a consequence of the Heine-Borel theorem, as the reader may readily show.

THEOREM III. *If the values of a continuous function $f(x)$ are known for a set of points everywhere dense in the interval $(a, b)$, then the value of $f(x)$ is determined for all points of $(a, b)$.*

Since $f(x)$ is continuous, it follows that the limit $\underset{x \to x_0}{L} f(x)$ exists at any point $x_0$ in $(a, b)$. The value of this limit is determined if we allow $x$ to approach $x_0$ through any set of points dense at $x_0$. Since the given function is known for a set of points everywhere dense in $(a, b)$, it follows that its value at each point $x_0$ of $(a, b)$ is given by the limit

$$\underset{x_n \to x_0}{L} f(x) = f(x_0),$$

where $\{x_n\}$ is any set dense at $x_0$.

Thus, if a given function is known to be continuous and its values are known for the rational points or the irrational points, its values are determined at every point. If a continuous function has the same value at a set of points everywhere dense, it is fully defined and is everywhere a constant.

THEOREM IV. *If $f(x)$ is single-valued and continuous in the closed interval $(a, b)$, then there exists a definite point $x_0$ of $(a, b)$ at which $f(x)$ takes its maximum value.*

As the given interval is closed and the function is continuous, it follows, as we have seen, that $f(x)$ is bounded and its least upper bound in $(a, b)$ is identical with its upper limit. Denote the upper limit of $f(x)$ by $M$. We are to show that there is a definite point $x_0$ of $(a, b)$ where

$$f(x_0) = M.$$

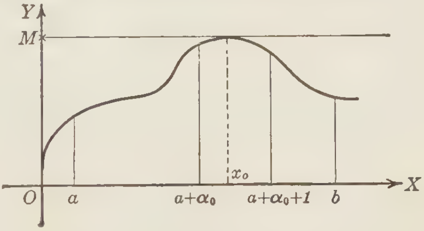


FIG. 19.

To do this, we divide the given interval $(a, b)$ into the subintervals

$$(a, a + 1),\ (a + 1, a + 2), \cdot \cdot \cdot , (a + n, b),$$

where all, with the exception of at most the last, are of unit length. In each of these subintervals $f(x)$ has an upper limit. None of these upper limits can exceed $M$, and there must be at least one subinterval in which it is equal to $M$. As $x$ increases from $x = a$, suppose the first interval encountered in which $f(x)$ has the upper limit $M$ is $(a + \alpha_0, a + \alpha_0 + 1)$. Divide this interval into 10 equal parts and in at least one of these subintervals, say

$$\left(a + \alpha_0 + \frac{\alpha_1}{10}, a + \alpha_0 + \frac{\alpha_1}{10} + \frac{1}{10}\right),$$

$f(x)$ has the upper limit $M$. Divide this interval likewise into 10 equal parts and proceed as before, continuing the process indefinitely. We obtain in this manner two auxiliary functions

$$\phi(n) = \alpha + \alpha_0 + \frac{\alpha_1}{10} + \frac{\alpha_2}{10^2} + \cdot \cdot \cdot + \frac{\alpha_n}{10^n},$$

$$\psi(n) = \alpha + \alpha_0 + \frac{\alpha_1}{10} + \frac{\alpha_2}{10^2} + \cdot \cdot \cdot + \frac{\alpha_n}{10^n} + \frac{1}{10^n}.$$

For values of $x$ such that

$$\phi(n) \leqq x \leqq \psi(n),$$

the function $f(x)$ has the upper limit $M$. As $n$ increases indefinitely $\phi(n)$ and $\psi(n)$ both approach a limit since both are bounded and monotone. Moreover, their limits are the same, since their difference $\frac{1}{10^n}$ approaches the limit zero. We may then write

$$\underset{n\to\infty}{L}\ \phi(n) = \underset{n\to\infty}{L}\ \psi(n) = x_0.$$

In other words, $f(x)$ approaches the upper limit $M$ as $x$ approaches $x_0$; that is,

$$\underset{x\to x_0}{L}\ f(x) = M.$$

But $f(x)$ is continuous at $x_0$ and hence

$$\underset{x\to x_0}{L}\ f(x) = f(x_0).$$

It follows from these two relations that

$$f(x_0) = M,$$

as the theorem requires. We may speak of this upper limit $M$ as the maximum value of the function since this value is also the least upper bound of the function and is a value of the function itself.

In a similar manner we may show that a function which is continuous in a closed interval takes on its minimum value at some point of the interval.

THEOREM V. *If $f(x)$ is a continuous function in the closed interval $(a, b)$, then as $x$ takes all values between $a$ and $b$, $f(x)$ takes at least once all values between $f(a)$ and $f(b)$.*

Without loss of generality we may assume

$$f(a) > f(b).$$

On the $X$-axis locate the points

$$a + 1,\ a + 2,\ a + 3,\ \cdot \cdot \cdot . \tag{1}$$

Let $y_0$ be any point on the $Y$-axis between $f(b)$ and $f(a)$. It is to be shown that for some definite value of $x$ the function $f(x)$ has the value $y_0$. If $y_0$ happens to be the maximum or the minimum value of $f(x)$ in any of the subintervals made by inserting the points (1), then by Theorem IV there exists a point $x_0$ of the given interval for which $f(x)$ takes the value $y_0$.

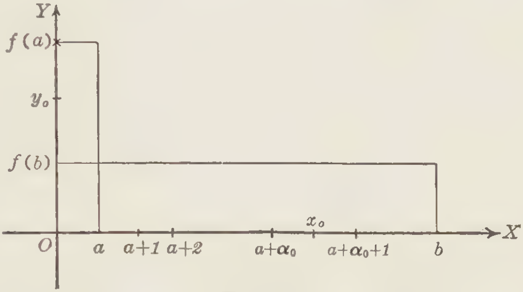


FIG. 20.

In case $y_0$ is not the maximum or the minimum value of $f(x)$ in any of the subintervals made by the insertion of the points (1), then in at least one of these subintervals, say $(a + \alpha_0, a + \alpha_0 + 1)$, there are values of $x$ for which $f(x)$ is larger than $y_0$ and also values for which it is smaller than $y_0$. Divide this interval into 10 equal parts. Again in some one of these subintervals, say

$$\left(a + \alpha_0 + \frac{\alpha_1}{10},\ a + \alpha_0 + \frac{\alpha_1}{10} + \frac{1}{10}\right),$$

there must be values of $x$ for which $f(x)$ is greater than $y_0$ and others for which $f(x)$ is smaller than $y_0$. Dividing this interval into 10 equal parts, we continue as before. We have in general the interval $[\phi(n), \psi(n)]$, where

$$\phi(n) = a + \alpha_0 + \frac{\alpha_1}{10} + \cdot\ \cdot\ \cdot + \frac{\alpha_n}{10^n},$$

$$\psi(n) = a + \alpha_0 + \frac{\alpha_1}{10} + \cdot\ \cdot\ \cdot + \frac{\alpha_n}{10^n} + \frac{1}{10^n},$$

which contains values of $x$ for which $f(x)$ is greater than $y_0$ and other values for which $f(x)$ is less than $y_0$. As $n$ takes all positive integral values, the resulting sequence of intervals defines in the limit a definite point $x_0$.

Because of the continuity of $f(x)$, we have

$$\underset{x\to x_0}{L} f(x) = f(x_0).$$

However, $y_0$ must lie between $f[\phi(n)]$ and $f[\psi(n)]$ or coincide with one of these values. As $n$ becomes infinite $x$ approaches $x_0$ and hence

$$\underset{n\to\infty}{L} f\{\phi(n)\} = \underset{n\to\infty}{L} f\{\psi(n)\} = \underset{x\to x_0}{L} f(x) = f(x_0). \qquad (2)$$

From the equality of the first two of these limits, we have from Theorem IV, Art. 24

$$\underset{n\to\infty}{L}\, f\{\phi(n)\} = \underset{n\to\infty}{L}\, f\{\psi(n)\} = y_0. \tag{3}$$

By comparison of the results in (2) and (3) we have finally

$$y_0 = f(x_0), \tag{4}$$

which establishes the theorem.

In a similar manner we can show that a continuous function takes all values between the maximum and the minimum values of the function in a given interval; that is, all values between its upper and its lower limits. The converse of this statement and of the foregoing theorem is not true; that is, it does not follow that a function is continuous in an interval $(a, b)$ because it takes all values between $f(a)$ and $f(b)$ or between the upper and the lower limiting values of the function in the interval. Thus the function

$$f(x) = \sin\frac{1}{x},\ x \neq 0,\ \text{where } f(0) = 0,$$

is discontinuous for $x = 0$ since $\underset{x\to 0}{L}\ \sin\frac{1}{x}$ does not exist. Nevertheless, $f(x)$ takes all values between its maximum and its minimum values in any interval inclosing the origin and indeed an infinite number of times.

In fact, there exist functions which take all intermediate values between its values at any two points and still are totally discontinuous in that interval.[1]

If we have given a function of two variables $f(x, y)$ the problem of continuity presents nothing new so long as we confine our attention to continuity with respect to each variable separately. The foregoing theorems may be extended to the case of continuity with respect to the two variables taken together. To do this one needs merely to substitute *region* for *interval* in the statement of the theorems and generalize the proof accordingly.

Continuity of $f(x, y)$ with respect to both variables at a given point $(x_0, y_0)$ implies the continuity at this point of the function with respect to $x$ and to $y$. It does not imply, however, that $f(x, y)$

---

[1] For example of such a function see Art. 27.

is continuous in each variable separately in the deleted neighborhood of this point. For example, let

$$\begin{aligned} f(x, y) &= x \sin \frac{1}{y}, \text{ for } x \neq 0,\ y \neq 0, \\ &= 0, \end{aligned}$$

for $x = 0$ and all values of $y$; also for $y = 0$ and all values of $x$. This function is fully defined in the neighborhood of $(0, 0)$, and is continuous in $(x, y)$ at the origin. However, it is not continuous in $y$ for $x \neq 0$, since the limit

$$\underset{y \to 0}{L} f(x_0, y) \equiv \underset{y \to 0}{L}\, x_0 \sin \frac{1}{y}, \ (x_0 \neq 0),$$

does not exist.

If $f(x, y)$ is known to be merely continuous in $x$ alone and in $y$ alone, then questions arise which involve the relations between continuity in the two variables and that in each variable separately. We shall now consider some of these questions.

THEOREM VI. *Let $f(x, y)$ be defined for the region $a \leqq x \leqq b$, $c \leqq y \leqq d$, in which it is continuous with respect to $x$ and to $y$. At every point $(x_0, y_0)$ of this region, we have*

$$\underset{x \to x_0}{L}\ \underset{y \to y_0}{L} f(x, y) = \underset{y \to y_0}{L}\ \underset{x \to x_0}{L} f(x, y) = f(x_0, y_0).$$

It will be seen that the function as thus defined satisfies the conditions for the interchange of iterated limits as given in Theorem VI, Art. 24; for, from the continuity with respect to $x$ and to $y$, we have condition (*a*) of that theorem, namely

$$\begin{aligned} \underset{x \to x_0}{L} f(x, y) &= f(x_0, y),\ y \neq y_0, \\ \underset{y \to y_0}{L} f(x, y) &= f(x, y_0),\ x \neq x_0. \end{aligned}$$

Since $f(x, y)$ is continuous in $y$ for all values of $x$, we have for $x = x_0$

$$\underset{y \to y_0}{L} f(x_0, y) = f(x_0, y_0),$$

which is condition (*b*) of the theorem quoted. We may show as follows that condition (*c*) is also satisfied. Because of the continuity in each of the variables, we have

$$|f(x_0, y_0) - f(x, y_0)| < \frac{\epsilon}{3}, \ |x - x_0| < \delta_{y_0}, \tag{6}$$

$$|f(x_0, y_n) - f(x_0, y_0)| < \frac{\epsilon}{3}, \ |y_n - y_0| < \eta, \tag{7}$$

$$|f(x, y_n) - f(x_0, y_n)| < \frac{\epsilon}{3}, \ |x - x_0| < \delta_{y_n},\ y_n \neq y_0. \tag{8}$$

Combining these inequalities, we have

$$|f(x, y_n) - f(x, y_0)| < \epsilon,$$

which holds for each value of $x$ in the smaller of the two intervals, $(x_0 - \delta_{y_0}, x_0 + \delta_{y_0})$, $(x_0 - \delta_{y_n} x_0 + \delta_{y_n})$, and for each value of $y_n$ sufficiently close to $y_0$. It therefore holds for any set $\{y_n\}$ dense at $y_0$. Since all three conditions are satisfied, the interchange of the order of limits is permissible and the theorem follows.

It is often necessary to examine the nature of the limiting function along the boundary of the region in which the function is defined. Suppose, for example, that $f(x, y)$ is continuous in $x$ in the region defined by the inequalities

$$a \leqq x \leqq b, \quad y_0 < y \leqq d.$$

For points on the boundary line $y = y_0$, let the function be defined by the limit

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0).$$

We shall examine the nature of the convergence of $f(x, y)$ to the limiting function $f(x, y_0)$. As will be seen, this problem is a generalization of the corresponding problem which arises in the convergence of an infinite series or of a sequence of functions.

Let $x_1$ be any constant value of $x$ in the interval $(a, b)$. We have then

$$\underset{y \to y_0}{L} f(x_1, y) = f(x_1, y_0),$$

which may be expressed as an inequality as follows

$$|f(x_1, y) - f(x_1, y_0)| < \epsilon, \quad |y - y_0| < \delta(x_1). \tag{9}$$

Geometrically, this relation means that for $x = x_1$ the values of $f(x_1, y)$ lie in the rectangle whose sides are given by the equations

$$z = f(x_1, y_0) \pm \epsilon, \quad y = y_0, \quad y = y_0 + \delta(x_1),$$

as indicated in Fig. 21.

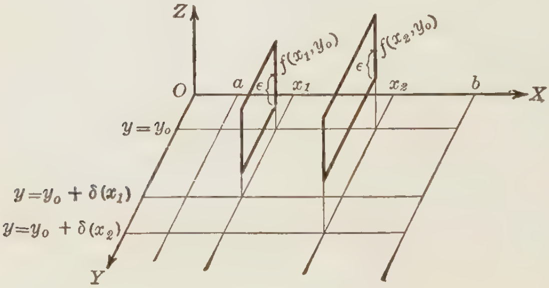


FIG. 21.

If we keep the value of $\epsilon$ fixed while the point $x$ is allowed to change, the value of $\delta(x)$ may vary with $x$. For example, for $x = x_2$ the value of $\delta(x_2)$ may be quite different from $\delta(x_1)$.

Thus, for each value of $x$ in the interval $(a, b)$ there is determined for a given $\epsilon$ a definite value of $\delta(x)$. This aggregate of values of $\delta(x)$ has a greatest lower bound $\delta$ in the interval $(a, b)$. If $\delta$ is different from zero, then $f(x, y)$ is said to **converge uniformly** to the limiting function $f(x, y_0)$. This kind of convergence is quite distinct from the convergence of $f(x, y)$ for each value of $x$ in the given interval. The latter means that for any value of $\epsilon$ there exists a $\delta(x)$ for each value of $x$ in $(a, b)$, which may, however, change with the point $x$. Uniform convergence, on the other hand, means that for a given value of $\epsilon$ there exists a value $\delta$ different from zero for which the inequality (9) holds uniformly for all values of $x$ in $(a, b)$; that is, $\delta$ is independent of $x$.

The curves whose equations are obtained from

$$z = f(x, y),$$

by putting $y = y_n$, where $y_n \neq y_0$, are called the **approximation curves** of the limiting function $z = f(x, y_0)$. These curves are those cut from the surface

$$z = f(x, y),$$

by the planes

$$y = y_n.$$

In the case of uniform convergence, these approximation curves become in the limit, as $y_n$ approaches $y_0$, the curve $z = f(x, y_0)$. The approximation curves can be conveniently exhibited by projecting them upon the $(XZ)$-plane. If the given function converges uniformly, then all of the projections of the approximation curves from a certain point on must fall within a strip of width $\epsilon$ on either side of the curve $z = f(x, y_0)$. The following example will illustrate a non-uniformly convergent function.

**Ex. 8.** Given $f(x, y) = \frac{xy}{x^2 + y^2}$, $0 \leqq x \leqq 1$, $0 < y \leqq 1$, where $f(x, 0) = \underset{y \to 0}{L} f(x, y)$. Draw three approximation curves and show that $f(x, y)$ does not converge uniformly to $f(x, 0)$.

The value of $f(x, 0)$ is zero for all values of $x$, and hence the curve

$$z = f(x, 0)$$

is the $X$-axis itself. For $y = y_n = \frac{1}{n}$, we have the approximation curves whose equation is

$$z = \frac{nx}{1 + n^2x^2}$$

The nature of the approximation curves is indicated in Fig. 22, each curve being projected upon the $(XZ)$-plane.

Each approximation curve has a peak of height $\frac{1}{2}$ for some value of $x$, no matter how large $n$ may be taken. As $n$ increases the peak in the curve grows sharper and approaches the origin, but it never disappears, nor does its height change. Hence, if we lay off a strip of width $\epsilon$ along the curve $z = f(x, 0)$, namely along the $X$-axis, the approximation curves never fall wholly within it. Consequently, the given function does not converge uniformly to the limiting function $f(x, 0)$, although the given function is everywhere continuous in $x$ and in $y$.

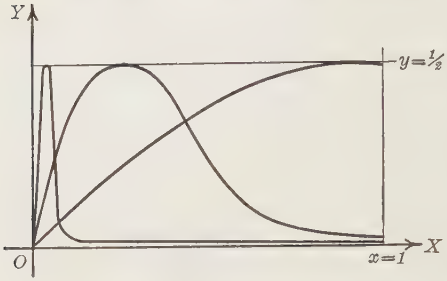


FIG. 22.

As a criterion for uniform convergence of $f(x, y)$ to the limiting function $f(x, y_0)$, we have the following theorem.

THEOREM VII. *Let* $f(x, y)$ *be continuous in* $x$ *for* $y \neq y_0$, $a \leqq x \leqq b$, *and defined for* $y = y_0$ *by the limit* $\underset{y\to y_0}{L} f(x, y) = f(x, y_0)$. *A necessary and sufficient condition that* $f(x, y)$ *converges uniformly to the limiting function* $f(x, y_0)$ *is that the simultaneous limit* $\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x, y)$ *exists at each point* $x_0$ *of the closed interval* $(a, b)$.

The condition of the theorem is necessary. For, let us assume that $f(x, y)$ converges uniformly, and let $x_0$ be any point of $(a, b)$. From the definition of $f(x, y_0)$, we have for a sufficiently small $\delta$

$$|f(x_0, y) - f(x_0, y_0)| < \epsilon, \quad |y - y_0| \leqq \delta. \tag{10}$$

Because of the continuity of $f(x, y)$ in $x$, we have on the line $y = y_0 + \delta$ an interval $(x_0 - \eta, x_0 + \eta)$, where $\eta$ is a function of $\delta$, such that we have for all values of $x$ in this interval

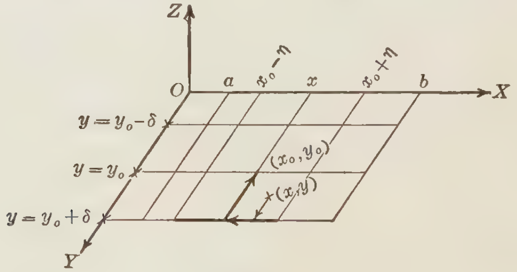


FIG. 23.

$$|f(x, y_0 + \delta) - f(x_0, y_0 + \delta)| < \epsilon.$$

By virtue of the uniform convergence of $f(x, y)$ we can select a $\delta$ independent of $x_0$ so that we have for all values of $x$ in $(x_0 - \eta, x_0 + \eta)$

$$|f(x, y) - f(x, y_0 + \delta)| < \epsilon.$$

The inequality (10) holds for $y = y_0 + \delta$, and hence we may write

$$|f(x_0, y_0 + \delta) - f(x_0, y_0)| < \epsilon.$$

Combining these three inequalities, we obtain

$$|f(x, y) - f(x_0, y_0)| < 3\epsilon, \ |y - y_0| < \delta, \ |x - x_0| < \eta; \quad (11)$$

that is, for a given $\epsilon$ there exists a region defined by the equation

$$y = y_0, \ y = y_0 + \delta, \ x = x_0 \pm \eta,$$

such that for any point $(x, y)$ in this region the value of $f(x, y)$ differs from $f(x_0, y_0)$ by less than $3\epsilon$. Hence, we have

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y) = f(x_0, y_0).$$

As $x_0$ was any point in the closed interval $(a, b)$, it follows that this limit exists at all points of the given interval.

The given conditions are also sufficient. For, we then have about each point $(x_0, y_0)$ of the closed interval $(a, b)$ a region defined by the inequalities

$$|x - x_0| < \delta, \ |y - y_0| < \delta,$$

such that for each point $(x, y)$ in this region we have

$$|f(x, y) - f(x_0, y_0)| < \epsilon, \quad (12)$$

where $\epsilon$ is arbitrarily chosen. It is to be understood that if $x_0$ is an extreme point of the given interval $(a, b)$, then only that portion of the above region is to be considered which corresponds to values of $x$ in $(a, b)$. Keeping $\epsilon$ fixed when once chosen, the value of $\delta$ depends upon the choice of $x_0$ and to each value of $x$ in the closed interval $(a, b)$ there is defined a definite value $\delta(x)$. The aggregate of these values has a greatest lower bound $\delta_1$.

This greatest lower bound $\delta_1$ is different from zero. For, assuming that it is zero there must be some point $x'$ of the closed interval $(a, b)$ in every neighborhood of which the greatest lower bound is zero. For $x = x'$, however, the value of $\delta$ must be different from zero, since the inequality (12) must hold for $x_0 = x'$, $y = y_0$. Put $\delta(x') = \delta'$. Then in the interval $\left(x' - \frac{\delta'}{2}, x' + \frac{\delta'}{2}\right)$ the value of $\delta(x)$ for each point is at least equal to $\frac{\delta'}{2}$. But, as $\delta'$ is different from zero, the greatest lower bound of $\delta(x)$ in this subinterval must be different from zero, which is contrary to our assumption. From this contradiction it follows that the conditions of the theorem are sufficient.

Another condition which is often convenient in testing the uniform convergence of a function is that given by the following theorem.

THEOREM VIII. *If, in the closed interval* $(a, b)$, $f(x, y_0)$ *is defined by the limit*

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0), \tag{a}$$

*and if for all values of* $y \neq y_0$ *and all values of* $x$ *in* $(a, b)$ *we have*

$$| f_x'(x, y) | < M, \tag{b}$$

*where* $M$ *is a finite number, then* $f(x, y)$ *converges uniformly to the limiting function* $f(x, y_0)$.

Any plane parallel to the $(XZ)$-plane cuts from the surface

$$z = f(x, y)$$

a curve. Let $\phi(x)$ be the angle which the tangent to this curve makes with the orthogonal projection of this tangent upon the $(X\,Y)$-plane. Then from condition $(b)$ of the theorem, we have

$$| f_x'(x, y) | = | \tan \phi\,(x) | < M$$

for all values of $x$ in $(a, b)$ and all values of $y$ different from $y_0$. For a given value of $\epsilon$ there then exists about any point $x_0$ of $(a, b)$ an interval $(x_0 - \delta, x_0 + \delta)$ within which we have

$$| f(x, y) - f(x_0, y) | < \epsilon,$$

where $y$ may have any value different from $y_0$.

From condition $(a)$ of the theorem, we have

$$| f(x_0, y) - f(x_0, y_0) | < \epsilon, \quad | y - y_0 | < \eta.$$

Combining these inequalities, we obtain

$$| f(x, y) - f(x_0, y_0) | < 2\epsilon, \quad | x - x_0 | < \delta, \quad | y - y_0 | < \eta,$$

which is the condition that

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y) = f(x_0, y_0)$$

for any point $x_0$ of the closed interval $(a, b)$. By Theorem VII it follows that $f(x, y)$ converges uniformly to the function $f(x, y_0)$.

**Ex. 9.** As $y$ approaches 0, test the uniform convergence of the function

$$f(x, y) = \arctan xy, \quad 0 \leqq x \leqq 1, 0 < y \leqq 1,$$

where
$$f(x, 0) \equiv \underset{y \to 0}{L} \arctan xy = 0.$$

Differentiating, we have

$$f_x'(x, y) = \frac{y}{1 + x^2y^2}.$$

For $0 \leqq x \leqq 1$, $0 < y \leqq 1$, we have

$$| f_x'(x, y) | = \left| \frac{y}{1 + x^2y^2} \right| < 1.$$

Hence, by the foregoing theorem the given function converges uniformly to the limiting function.

While this theorem gives us a convenient test for the uniform convergence of a function, it does not give a test for the non-uniform convergence of a function; for, it is a sufficient but not a necessary test, as the following exercise shows.

**Ex. 10.** Test for uniform convergence the function

$$f(x, y) = \sqrt{xy}, \quad 0 \leqq x \leqq 1, \quad 0 < y \leqq 1,$$

where
$$f(x, 0) \equiv \underset{y \to 0}{L} \sqrt{xy} = 0.$$

We have for $x \neq 0$, $y \neq 0$

$$f_x'(x, y) = \frac{1}{2}\sqrt{\frac{y}{x}},$$

which has no finite upper bound in the neighborhood of $x = 0$, $y \neq 0$. The conditions of the foregoing theorem are not satisfied. However, the function converges uniformly; for,

$$\underset{\substack{x \to x_0 \\ y \to 0}}{L} f(x, y) = \underset{\substack{x \to x_0 \\ y \to 0}}{L} \sqrt{xy} = 0 = f(x_0, 0)$$

for every value of $x_0$ in the closed interval (0, 1); that is, the conditions of Theorem VII, which are both necessary and sufficient for uniform convergence, are satisfied.

A useful criterion for continuity of the limiting function $f(x, y_0)$ is given by the following theorem.

THEOREM IX. *Let $f(x, y)$ be continuous in $x$ for $y \neq y_0$ and let it converge uniformly to the limiting function $f(x, y_0)$ in the interval $(a, b)$. Then $f(x, y_0)$ is continuous in $x$.*

Let $x_0$ be any point in $(a, b)$. Since $f(x, y)$ converges uniformly to the function $f(x, y_0)$, we have by Theorem VII

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y) = f(x_0, y_0).$$

Moreover, as $f(x, y)$ is continuous in $x$ for $y \neq y_0$, we have

$$\underset{x \to x_0}{L} f(x, y) = f(x_0, y).$$

From the definition of $f(x, y_0)$ we have

$$\underset{y \to y_0}{L} f(x_0, y) = f(x_0, y_0),$$

or combining these results, we may write

$$\underset{y \to y_0}{L} \underset{x \to x_0}{L} f(x, y) = f(x_0, y_0).$$

The conditions of Theorem V, Art. 24 are satisfied and hence we may interchange the order of limits and write

$$\underset{x \to x_0}{L} \underset{y \to y_0}{L} f(x, y) = \underset{x \to x_0}{L} f(x, y_0) = f(x_0, y_0).$$

From this last equality it follows that $f(x, y_0)$ is continuous at $x_0$. As $x_0$ is any point of the closed interval $(a, b)$ it follows that $f(x, y_0)$ is continuous throughout the interval.

That the foregoing condition is sufficient but not necessary for the continuity of $f(x, y_0)$ is at once apparent. For example, the function given in Ex. 8 converges to a continuous function of $x$, but the convergence is non-uniform.

A necessary and sufficient condition that $f(x, y_0)$ is continuous at any point $x_0$ of the given interval may be stated as follows.[1]

THEOREM X. *Let $f(x, y)$ be continuous in $x$ for $y \neq y_0$, and let $f(x, y_0) = \underset{y \to y_0}{L} f(x, y)$ for each value of $x$ in a closed interval $(a, b)$. A necessary and sufficient condition that $f(x, y_0)$ is continuous in $x$ at $x = x_0$, $a \leqq x_0 \leqq b$, is that for an arbitrarily small positive $\epsilon$ and for a set of values $\{y_n\}$ dense at $y_0$, there exists an interval $(x_0 - \delta_{y_n}, x_0 + \delta_{y_n})$, whose length is different from zero and may vary with $y_n$, such that for all values of $x$ in this interval we have*

$$| f(x, y_n) - f(x, y_0) | < \epsilon.$$

This theorem follows at once from the condition given in Theorem VI of Art. 24 for the interchange of the order of limits. We have given the condition that

$$\underset{y \to y_0}{L} \underset{x \to x_0}{L} f(x, y) = \underset{y \to y_0}{L} f(x_0, y) = f(x_0, y_0).$$

---

[1] See Arzelà, *Sulle serie di Funzioni* (1899) Parti Prima, p. 14.

By the interchange of limits this relation gives

$$\underset{x \to x_0}{L} \; \underset{y \to y_0}{L} f(x, y) = \underset{x \to x_0}{L} f(x, y_0) = f(x_0, y_0),$$

and hence $f(x, y_0)$ is continuous at $x = x_0$.

The foregoing theorem gives a necessary and sufficient condition that $f(x, y_0)$ is continuous at a single point. We may generalize this result so as to get a condition for the continuity of $f(x, y_0)$ throughout an interval by introducing what we shall call **quasi-uniform convergence,**[1] which may be defined as follows.

As in the foregoing discussion, let $f(x, y_0)$ be defined by the relation

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0)$$

for each value of $x$ in the closed interval $(a, b)$. The function $f(x, y)$ is said to converge quasi-uniformly to the limiting function $f(x, y_0)$ in the interval $(a, b)$, if for an arbitrarily small positive number $\epsilon$ and for some $y_n \neq y_0$, there exists a positive number $\delta$ and a finite set of values of $y$, say

$$y_1, \; y_2, \; y_3, \; \cdot \; \cdot \; \cdot \; , y_k, \; \cdot \; \cdot \; \cdot \; , y_{n-1},$$

lying between $y_0$ and the given value $y_n$ such that on the lines

$$y = y_k, \; k = 1, \; 2, \; \cdot \; \cdot \; \cdot \; , n - 1,$$

it is possible to choose a finite number of segments, each of length at least equal to $\delta$ such that the sum of their projections on $y = y_0$ completely covers the interval $(a, b)$ and furthermore for all values of $x$ in this interval we have

$$|f(x, y_0) - f(x, y_k)| < \epsilon, \quad (13)$$

the point $(x, y_k)$ being a point on some one of these segments.

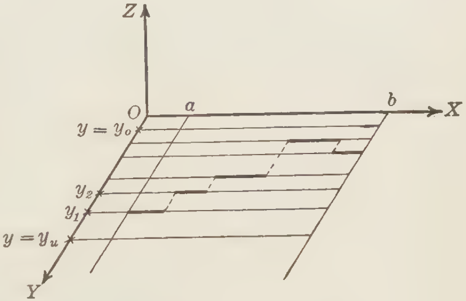


FIG. 24.

Quasi-uniform convergence differs from ordinary uniform convergence in that in the latter case the condition (13) must hold not only for all values of $x$ but for all values of $y$ in the neighborhood of $y_0$. If a given function converges uniformly to its limiting function, it also converges quasi-uniformly. The converse, however, is not true.

---

[1] Called by Arzelà " uniform convergence by segments." The term "quasi-uniform convergence " is due to Borel. See *Encyclop. des sci. math.* $II_2$, Art. 52. Quasi-uniform convergence is the more descriptive term, since there are functions satisfying the Arzelà conditions for convergence which do not converge uniformly in any subinterval.

If the given function $f(x, y)$ satisfies the foregoing conditions for quasi-uniform convergence except for a finite set of subintervals, the sum of whose lengths is arbitrarily small, then we say that $f(x, y)$ converges to the limiting function $f(x, y_0)$ **quasi-uniformly in general,** or quasi-uniformly almost everywhere.

We may now state the following theorem.

THEOREM XI. *If for $y \neq y_0$ the function $f(x, y)$ is continuous in $x$ in the closed interval $(a, b)$, a necessary and sufficient condition that the limiting function $f(x, y_0)$ also is a continuous function of $x$ in $(a, b)$ is that $f(x, y)$ converges quasi-uniformly to $f(x, y_0)$ in this interval.*

From the continuity in $x$ of $f(x, y)$ for $y \neq y_0$ and from the definition of the limiting function, it follows that

$$\underset{y \to y_0}{L} \; \underset{x \to x_0}{L} \; f(x, y) = \underset{y \to y_0}{L} \; f(x_0, y) = f(x_0, y_0), \tag{14}$$

where $x_0$ is any point in the closed interval $(a, b)$. Because $f(x, y)$ converges quasi-uniformly to the limiting function $f(x, y_0)$, the conditions for the interchange of the order of iterated limits given in Theorem VI, Art. 24 are satisfied for any point $x_0$ of $(a, b)$. Hence we have from (14)

$$\underset{x \to x_0}{L} \; \underset{y \to y_0}{L} \; f(x, y) = \underset{x \to x_0}{L} \; f(x, y_0) = f(x_0, y_0).$$

Since $x_0$ is any point in $(a, b)$ it follows from the existence of this last limit that $f(x, y_0)$ is a continuous function of $x$ in the closed interval $(a, b)$.

In quasi-uniform convergence, an essential condition is that there shall exist a set of values $\{y_n\}$ dense at $y_0$. In the foregoing discussion, we have assumed that $f(x, y_n)$ is continuous in $x$. If we replace the sequence $\{y_n\}$ by the sequence $\{n\}$ of positive integers, where $n$ may be regarded as a subscript, the foregoing theorems give at once important results concerning the continuity of the limiting function $f(x)$ of a sequence $\{f_n(x)\}$ of continuous functions. For example, Theorem X gives a necessary and sufficient condition that $f(x)$ is continuous at a given point $x_0$. It follows from the theorems on uniform convergence that, if the sequence of continuous functions converges uniformly in the interval $(a, b)$, then $f(x)$ is continuous throughout that interval. The condition of uniform convergence, however, is sufficient but not necessary for continuity. Both a necessary and sufficient condition for the continuity of $f(x)$ is given by quasi-uniform convergence of the given sequence $\{f_n(x)\}$. As we shall see later, these conclusions apply also to the character of a function defined by a convergent infinite series, since its convergence may be made to depend upon the convergence of an infinite sequence.

**27. Semi-continuous functions.** As we have seen, the maximum value of $f(x)$ at a point $x_0$, namely $M(f, x_0)$, is the largest of the three numbers $f(x_0)$, $\overline{f(x_0 + 0)}$, $\overline{f(x_0 - 0)}$. Let us now suppose that

$$f(x_0) = M(f, x_0).$$

If, for an arbitrarily small positive $\epsilon$, there exists a $\delta > 0$, such that for all values of $x$ in the neighborhood $(x_0 - \delta, x_0 + \delta)$ we have

$$f(x) < f(x_0) + \epsilon, \tag{1}$$

then $f(x)$ is said to be **upper semi-continuous** at $x_0$, or $x_0$ is a point of upper semi-continuity of $f(x)$.

Similarly, if we have
$$f(x_0) = m(f, x_0),$$

where $m(f, x_0)$ is the minimum value of $f(x)$ at $x_0$, then we say that $f(x)$ is **lower semi-continuous** at $x_0$, if there exists a neighborhood $(x_0 - \delta, x_0 + \delta)$ such that for all values of $x$ in this interval

$$f(x) > f(x_0) - \epsilon. \tag{2}$$

One may also characterize semi-continuous functions by their relation to continuous functions. In order that $f(x)$ shall be continuous at $x_0$, one must have

$$|f(x) - f(x_0)| < \epsilon, \quad |x - x_0| < \delta,$$

which is equivalent to the two conditions

$$f(x) < f(x_0) + \epsilon, \quad |x - x_0| < \delta,$$
$$f(x) > f(x_0) - \epsilon, \quad |x - x_0| < \delta.$$

Upper semi-continuity implies that the first of these conditions holds, while the second may not hold, and also that $f(x_0)$ is the maximum value of $f(x)$ at the point. A similar statement may be made concerning lower semi-continuity at $x_0$.

A function is **upper semi-continuous throughout an interval** if each point of the interval is a point of upper semi-continuity. Similarly, a function is **lower semi-continuous in an interval** provided each point of the interval is a point of lower semi-continuity. If a function is both upper semi-continuous and lower semi-continuous, it is continuous, and conversely if it is continuous it has both of these properties.

As an illustration of an upper semi-continuous function, we have the function $M(f,x)$ discussed in Art. 25. Similarly the function $m(f, x)$ is an illustration of a lower semi-continuous function.

As another illustration of an upper semi-continuous function, we may consider the values of $\delta$ defined in connection with continuous functions. From the continuity of a function $f(x)$, we have for any $x_0$ in a given interval $(a, b)$,

$$|f(x) - f(x_0)| < \eta, \; |x - x_0| < \delta, \tag{3}$$

where $\eta$ is arbitrarily small and $\delta$ depends upon $\eta$ and upon the choice of $x_0$. For any given $\eta$ there is defined a $\delta(x)$ for each $x$ in the given interval $(a, b)$. It is this function $\delta(x)$ that we shall consider, it being assumed that $\delta(x)$ is always taken as large as possible to satisfy the conditions given in (3). If $f(x)$ is constant in $(a, b)$, it follows at once that $\delta(x)$ is upper semi-continuous; for, it is the distance from $x_0$ to the nearest extremity of the interval $(a, b)$ and in fact in this case $\delta(x)$ varies continuously with $x$.

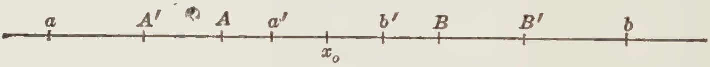


FIG. 25.

We shall now consider the case where $f(x)$ is not a constant in the neighborhood of $x_0$. Let $A$, $B$ be the extremities of the interval $\{x_0 - \delta(x_0), x_0 + \delta(x_0)\}$. For a proper choice of $\eta$ we can select some value $x_1$ in the neighborhood of $x_0$ but distinct from it, such that one extremity of the corresponding interval $[x_1 - \delta(x_1), x_1 + \delta(x_1)]$ will fall outside of $AB$, while the other extremity may lie within $AB$. This statement is true irrespective of the side of $x_0$ on which $x_1$ is situated, due to the fact that $f(x)$ is continuous and the value of $\delta(x)$ is always chosen as large as possible. Allow $x$ to range over such values in the neighborhood of $x_0$ that the extremities of the corresponding interval $\{x - \delta(x), x + \delta(x)\}$ satisfy the foregoing condition that at least one falls exterior to $AB$ and the other lies on the opposite side of $x_0$. Denote the interval within which the extremities of the $\delta(x)$ interval may lie by $A'$, $B'$. By choosing the points $A'$

and $B'$ arbitrarily, one restricts the range within which $x_1$ must lie. For example, we may put

$$A' = x_0 - \delta(x_0) - \epsilon, \quad B' = x_0 + \delta(x_0) + \epsilon,$$

where $\epsilon$ is arbitrarily small. Denote by $(a'b')$ the corresponding neighborhood of $x_0$ over which $x_1$ may range and neither extremity of the interval $[x_1 - \delta(x_1), x_1 + \delta(x_1)]$ lie exterior to $A'B'$. Then for all values of $x$ within $(a'b')$, we have

$$\delta(x) < \delta(x_0) + \epsilon;$$

that is, $\delta(x)$ satisfies the definition for an upper semi-continuous function at the point $x_0$. But as $x_0$ is any point in the given interval $(a, b)$, it follows that $\delta(x)$ is upper semi-continuous throughout the interval. In case $f(x)$ is uniformly continuous in $(a, b)$ it follows that the upper semi-continuous function $\delta(x)$ has a positive lower bound. If, on the other hand, $f(x)$ is continuous but not uniformly so, then $\delta(x)$ has zero as its greatest lower bound.

THEOREM I. *Given a finite number of functions which have a common point of upper semi-continuity, then their sum is also upper semi-continuous at that point.*

It is sufficient to prove the theorem for the sum of two upper semi-continuous functions. The same method applies to the sum of any finite number of such functions. A similar theorem may be established for lower semi-continuous functions.

Let the two given functions be $\phi_1(x)$ and $\phi_2(x)$, and let $x_0$ be a point at which both are upper semi-continuous. We then have for all values of $x$ in some interval $(x_0 - \delta_1, x_0 + \delta_1)$ the relation

$$\phi_1(x) < \phi_1(x_0) + \frac{\epsilon}{2}.$$

Similarly, there exists an interval $(x_0 - \delta_2, x_0 + \delta_2)$ such that for all values of $x$ within it, we have

$$\phi_2(x) < \phi_2(x_0) + \frac{\epsilon}{2}.$$

Combining these two inequalities, we have

$$\phi_1(x) + \phi_2(x) < \phi_1(x_0) + \phi_2(x_0) + \epsilon,$$

which holds for all values of $x$ within $(x_0 - \delta, x_0 + \delta)$, where this interval is the smaller of the two intervals

$$(x_0 - \delta_1, x_0 + \delta_1), \quad (x_0 - \delta_2, x_0 + \delta_2).$$

Hence, the sum of the given functions is upper semi-continuous, as the theorem requires.

THEOREM II. *If $f(x)$ is an upper semi-continuous function at $x = x_0$ then $-f(x)$ is a lower semi-continuous function at that point.*

Because $f(x)$ is upper semi-continuous at any point $x_0$, we have

$$f(x) < f(x_0) + \epsilon, \; x_0 - \delta < x < x_0 + \delta.$$

Changing the signs of each member of this inequality, we have

$$-f(x) > -f(x_0) - \epsilon;$$

that is, $-f(x)$ is a lower semi-continuous function at $x_0$.

Similarly, it may be shown that if $f(x)$ is a lower semi-continuous function at $x_0$, then $-f(x)$ is an upper semi-continuous function at the same point.

THEOREM III. *The saltus function $\omega(f, x)$ is an upper semi-continuous function.*

The saltus function was defined by the relation

$$\omega(f, x) = M(f, x) - m(f, x),$$

where $M(f, x)$ denotes the maximum value and $m(f, x)$, the minimum value of $f(x)$ at each point $x$ of a given interval. As we have seen, $M(f, x)$ is an upper semi-continuous function. By Theorem II it follows that $-m(f, x)$ is an upper semi-continuous function. The function $\omega(f, x)$ can then be written as the sum of two upper semi-continuous functions as follows

$$\omega(f, x) = M(f, x) + \{-m(f, x)\}.$$

It is therefore an upper semi-continuous function, as the theorem states.

In a similar manner, it may be shown that the saltus function of a function of two independent variables is also an upper semi-continuous function.

THEOREM IV. *If $f(x)$ is an upper semi-continuous function in the interval $(a, b)$, then the points of $(a, b)$ where*

$$f(x) \geqq A,$$

*$A$ being any arbitrarily chosen number, form a closed set.*

Let $x_0$ be a limiting point of a sequence $\{x_n\}$ at each point of which we have

$$f(x_n) \geqq A.$$

In every neighborhood $(x_0 - \delta, x_0 + \delta)$ of $x_0$ the least upper bound $M(f, \delta)$ of $f(x)$ for all values of $x$ is at least equal to the least upper bound of $f(x_n)$. As $\delta$ approaches zero, the limit of $M(f, \delta)$ is the maximum value of $f(x)$ at the point $x_0$. Hence, we have

$$M(f, x_0) \geqq A; \tag{4}$$

for, as $x$ takes the values of the set $\{x_n\}$, the corresponding values of $f(x)$ are all equal to or greater than $A$. As $f(x)$ is an upper semi-continuous function at $x_0$, then

$$M(f, x_0) = f(x_0),$$

whence from (4)

$$f(x_0) \geqq A.$$

In a similar manner it can be shown that the points at which a lower semi-continuous function is less than a given number form by themselves a closed set.

The foregoing theorem can be extended to functions of two or more variables. The proofs are similar, the functions $\omega(f, x)$, $M(f, x)$, $m(f, x)$ being replaced respectively by the corresponding functions for $f(x, y)$, namely $\omega(f, x, y)$, $M(f, x, y)$, $m(f, x, y)$.

**28. Discontinuous functions.** In this article we shall consider particularly those functions having an infinite number of points of discontinuity. Such functions are totally discontinuous if every point is a point of discontinuity or they may be point-wise discontinuous if the points of continuity form a set everywhere dense. Theorem III of the last article states that the saltus function is upper semi-continuous and by Theorem IV of the same article the points where such a function exceeds an arbitrarily chosen number form a closed set. This leads to the conclusion that the set of discontinuities where the amount of the discontinuity exceeds an arbitrarily chosen number is necessarily closed. If these points are everywhere dense in any interval, then every point is such a point. Consequently, totally discontinuous functions are further characterized by the condition that the points where the saltus is greater than an arbitrarily selected $A$ form a set everywhere dense.

THEOREM I. *A semi-continuous function is at most point-wise discontinuous.*

Let $f(x)$ be any function defined for the interval $(a, b)$, and for convenience let

$$F(x) \equiv M(f, x) - f(x). \tag{1}$$

Consequently, $F(x)$ is either positive or zero at each point of $(a, b)$. The function $M(f, x)$ is upper semi-continuous in the same interval. Denote by $(a', b')$ any subinterval of $(a, b)$, and let $x_0$ be any point of $(a', b')$. Because $M(f, x)$ is upper semi-continuous at $x_0$, we have for all values of $x$ in the neighborhood of $x_0$

$$M(f, x) < M(f, x_0) + \epsilon. \tag{2}$$

Since $M(f, x_0)$ is also the maximum value of $f(x)$ at $x_0$, it follows that in every neighborhood of $x_0$ a point $x_1$ can be found for which

$$f(x_1) > M(f, x_0) - \epsilon. \tag{3}$$

The inequality (2) holds for all values of $x$ in every neighborhood of $x_0$ and hence for $x = x_1$. We have then from (2) and (3)

$$M(f, x_1) - f(x_1) < 2\epsilon.$$

As $\epsilon$ is an arbitrarily small number, it follows that $x_1$ is a point at which $F(x)$ takes the value zero. As $x_0$ is a point in $(a', b')$ and in every neighborhood of each such point there exists a point $x_1$, it follows that the points at which $F(x)$ has the value zero are everywhere dense in $(a, b)$.

Similarly, it may be shown that the points at which

$$f(x) - m(f, x)$$

has the value zero form likewise a set everywhere dense in $(a, b)$.

Let now $f(x)$ be an upper semi-continuous function; then we have

$$f(x) = M(f, x).$$

The saltus function is

$$\omega(f, x) \equiv M(f, x) - m(f, x) = f(x) - m(f, x). \tag{4}$$

But, as we have seen, the points are everywhere dense in $(a, b)$ where $f(x)$ and $m(f, x)$ are identical. It follows from (4) that at these points the saltus function is zero and hence $f(x)$ is continuous. Consequently, every upper semi-continuous function is at most point-wise discontinuous.

Similarly, one may show that every lower semi-continuous function is at most point-wise discontinuous. Hence, the theorem follows.

The foregoing theorem can be extended to functions of two or more variables. The proof is similar in that the interval is in that case replaced by a region.

THEOREM II. *The points at which $f(x)$ has a discontinuity equal to or greater than a given number $A$ form a closed set.*

This theorem follows at once as a consequence of Theorem IV of the last article. For, as we have seen, the saltus function $\omega(f, x)$ is an upper semi-continuous function and by the theorem quoted the points at which such a function has values greater than $A$ form a closed set.

It follows from the foregoing theorem that, if $f(x)$ has a discontinuity greater than $A$ at a set of points dense at some point, then that point itself is a point of discontinuity where the amount of discontinuity is equal to or greater than $A$.

THEOREM III. *A necessary and sufficient condition that a function $f(x)$ is point-wise discontinuous in a given interval $(a, b)$ is that the points at which the saltus exceeds an arbitrarily chosen number $A$ do not form a set everywhere dense in any subinterval of $(a, b)$.*

The condition is necessary; for, by Theorem II, the points in question form a closed set and, if they are also everywhere dense in any interval, then in that subinterval every point is a point of discontinuity and $f(x)$ can not be point-wise discontinuous.

The condition stated in the theorem is also sufficient. To show this, we consider any sequence of numbers

$$\eta_1 > \eta_2 > \eta_3 > \cdot \cdot \cdot > \eta_n > \cdot \cdot \cdot$$

such that

$$\underset{n\to\infty}{L} \eta_n = 0.$$

By the conditions of the theorem the set of points at which the discontinuity is equal to or greater than $\eta_1$ can not be everywhere dense in any subinterval $(a_1, b_1)$ of $(a, b)$. There must then be some portion of $(a_1, b_1)$ which is free from such points. Let $(a_2, b_2)$ be such a subinterval of $(a_1, b_1)$. Likewise, the points where the discontinuity is equal to or greater than $\eta_2$ can not be everywhere dense in $(a_2, b_2)$ and there must be a subinterval $(a_3, b_3)$ of it which is free of such points. In the same way we can find a subinterval $(a_4, b_4)$ of $(a_3, b_3)$ which is free of points where the discontinuity is greater than $\eta_3$, etc. In general we find in this manner a subinterval $(a_n, b_n)$ free from points where the discontinuity is greater than $\eta_{n-1}$. If the process is continued indefinitely, we thus obtain a sequence of intervals

$$(a_1, b_1),\ (a_2, b_2),\ (a_3, b_3),\ \cdot \cdot \cdot,\ (a_n, b_n),\ \cdot \cdot \cdot$$

each lying within the previous interval. If we now impose the further condition that

$$\underset{n\to\infty}{\mathrm{L}} (a_n - b_n) = 0,$$

this sequence of intervals defines a definite point $x_0$. At this point the discontinuity of $f(x)$ can not exceed the limit of $\eta_n$ as $n$ becomes infinite, and as the amount of the discontinuity, that is the saltus, is never negative it must be zero. In other words, $x_0$ is a point of continuity of $f(x)$. As $(a_1, b_1)$ was any subinterval of $(a, b)$, it follows that under the conditions of the theorem the given function must have at least one point of continuity in every subinterval and is therefore at most point-wise discontinuous.

It follows that the most general set of points that we can have where the discontinuity of a point-wise discontinuous function exceeds a given number $A$ is a perfect set which is in no subinterval everywhere dense, or such a set plus a set which is in no part of $(a, b)$ everywhere dense. It is not to be understood that the set of all points of discontinuity of a point-wise discontinuous function may not be everywhere dense in the interval of definition. Indeed such may be the case, as we shall see later, provided the saltus function has the lower limit zero at a set of points which is everywhere dense.

It follows as a consequence of this theorem that the sum or product of any finite number of point-wise discontinuous functions can not be totally discontinuous; for, the points where the discontinuity exceeds $A$ in the sum or product can not exceed the sum of the sets of points where the individual functions have a discontinuity greater than $A$, and this sum can not be everywhere dense.

THEOREM IV. *The points of continuity of a point-wise discontinuous function form a set having the cardinal number of the continuum.*

As in the proof of Theorem III, let

$$\eta_1 > \eta_2 > \eta_3 > \cdot \cdot \cdot > \eta_n > \cdot \cdot \cdot$$

be a sequence of numbers such that

$$\underset{n \to \infty}{L} \eta_n = 0.$$

Let $E_1$ be the set of points where the saltus of $f(x)$ is equal to or greater than $\eta_1$, $E_2$, the set of points where it is equal to or greater than $\eta_2$, etc., $E_n$ being that set of points where $f(x)$ has a saltus equal to or greater than $\eta_n$. Each set in the sequence of sets

$$E_1,\ E_2,\ \cdot \cdot \cdot,\ E_n,\ \cdot \cdot \cdot$$

is in no portion of the given interval $(a, b)$ everywhere dense. From the definition of a set of points of the first category, it follows that the sum of the sequence $\{E_n\}$ is a set $E$ of the first category; moreover, it is the set of all of the points of discontinuity of $f(x)$. The complementary set $C(E)$ is the set composed of all the points of continuity of $f(x)$. As $E$ is of the first category $C(E)$ must be of the second category. By Theorem V of Art. 18, $C(E)$, that is the set composed of the points of continuity of $f(x)$, must therefore have the cardinal number of the continuum.

It is a consequence of this theorem that a point-wise discontinuous function can not be discontinuous at the irrational points; for, the complementary set, namely, the set of all rational numbers of the given interval, could not serve as the points of continuity of such a function since the set of rational points is enumerable and hence can not have the cardinal number of the continuum. In fact it is impossible to have a function which is discontinuous at the irrational points and continuous at the rational points. For, in that case the points of continuity would be everywhere dense in the given interval and the function would be at most point-wise discontinuous but, as we have just seen, such a function must be continuous at a set of points having the cardinal number of the continuum. In fact it has been shown that in an interval in which a function $f(x)$ is point-wise discontinuous there does not exist another function which is continuous where $f(x)$ is discontinuous and conversely.[1]

Point-wise discontinuity and total discontinuity describe properties of a function with reference to its behavior with respect to an interval or at least to a set of points which is dense in itself. The characteristic property which distinguishes the one class of functions from the other is the distribution in the given interval of the set of points where the saltus of the function is equal to or greater than an arbitrarily chosen number $A$. If this set is not everywhere dense in any portion of the given interval the function is point-wise discontinuous; if it is everywhere dense throughout the interval, it is totally discontinuous.[2]

There are discontinuous functions which in a given interval are neither point-wise discontinuous nor totally discontinuous, as we have defined these terms. For example, in the interval $(-1, 1)$ let

---

[1] See Volterra, *Giorn. di mat.* Vol. 19, p. 76.

[2] Totally discontinuous functions are sometimes defined as those discontinuous functions which are not point-wise discontinuous. See Baire, *Leçons sur les fonctions discontinues*, p. 75; Caratheodory, *Vorlesungen über Reelle Funktionen*, p. 148.

$$\begin{aligned} f(x) &= x^2, \text{ for irrational values of } x, \\ &= 0, \text{ for rational values of } x. \end{aligned}$$

This function is discontinuous at every point except $x = 0$. In the given interval, $f(x)$ is then neither point-wise discontinuous nor totally discontinuous.

Instead of being continuous at a single point, a non-point-wise discontinuous function $f(x)$ may be continuous at a set of points, but this set of points can not be everywhere dense in a given interval. The points of discontinuity of such a function must be everywhere dense in the given interval but they can not form a closed set. A function having these two properties may be called a **point-wise continuous function,** in distinction from a point-wise discontinuous function on the one hand and from a totally discontinuous function on the other. It may be shown that the points of continuity form an **inner limiting set;**[1] that is, a set of points such that each is an inner point of at least one interval of every set of intervals selected from a given sequence of sets of intervals. The points of continuity of a point-wise continuous function may be enumerable or may even form a perfect set and hence have the cardinal number of the continuum. For, as we have seen (Art. 17), it is possible to have a perfect set which is in no subinterval everywhere dense.

The following theorem often gives a convenient test for a totally discontinuous function.

THEOREM V. *A function which is defined for a given interval* $(a, b)$ *and takes the same two distinct values in every arbitrarily small subinterval is totally discontinuous in* $(a, b)$.

Denote by $A$ and $B$ the two constant values which $f(x)$ takes in every arbitrarily small subinterval of $(a, b)$. If $x_0$ is any point of $(a, b)$ then there must exist in the neighborhood of $x_0$ a set of points $\{x_n\}$ dense at $x_0$ such that as $x$ is allowed to approach $x_0$ through this set we have

$$\underset{x_n \to x_0}{L} f(x_n) = A.$$

There must likewise exist an infinite set $\{x'_n\}$ of points dense at $x_0$ at which $f(x)$ takes the value $B$, and hence we may write

$$\underset{x'_n \to x_0}{L} f(x'_n) = B.$$

---

[1] See Young, *Sitzgsb, Akad. Wien* Vol. 112 (1903) Part 2a, p. 1307; also Schoenflies, *ibid.* (1904), Part 2a, p. 1277. For definition of an inner limiting set and its properties see Young's *Theory of Sets of Points*, p. 63, *et seq.*

But as $A$ is different from $B$ it follows that the limit $\underset{x\to x_0}{L} f(x)$ does not exist, and hence $f(x)$ is discontinuous at $x_0$. Since $x_0$ is any point of $(a, b)$ it follows that $f(x)$ is totally discontinuous in the given interval.

In Art. 26 it was pointed out that a function might take all values between any two of its values and still be totally discontinuous. The following example furnishes an illustration of such a function.

**Ex.**[1] **1.** Show that $f(x)$, defined as follows for the interval $(0, 1)$, takes all values intermediate between any two of its values and yet is totally discontinuous. The variable $x$ is expressed as a decimal fraction,

$$x = 0.x_1x_2x_3 \cdot \cdot \cdot = \frac{x_1}{10} + \frac{x_2}{10^2} + \frac{x_3}{10^3} + \cdot \cdot \cdot .$$

If the digits in the odd places, namely $x_1, x_3, x_5 \cdot \cdot \cdot$, form a non-periodic sequence, let $f(x)$ be zero. If these digits do form a periodic sequence from some point on, say after $x_{2n-1}$, let $f(x)$ be defined by the relation

$$f(x) = \frac{x_{2n}}{10} + \frac{x_{2n+2}}{10^2} + \frac{x_{2n+4}}{10^3} + \cdot \cdot \cdot .$$

The function $f(x)$ is fully defined for all values of $x$ in the interval $(0, 1)$, since for each value of $x$ there is a definite value of $f(x)$. Let $(a, b)$ be any arbitrarily small subinterval of $(0, 1)$. We shall show that $f(x)$ takes all the values between zero and one as $x$ varies from $a$ to $b$. It will then follow that $f(x)$ is totally discontinuous in $(0, 1)$ by the foregoing theorem; for, since it takes all values between zero and one, it can be said to take any two of these values in every such subinterval.

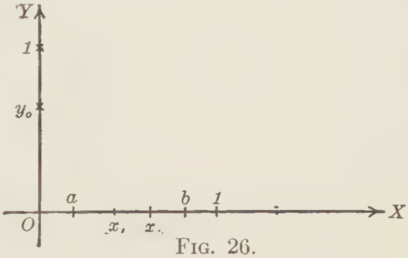


FIG. 26.

Let $y_0$ be any point on the $Y$-axis in the interval $(0, 1)$. We shall show that we can determine a point $x_0$ in the subinterval $(a, b)$ for which

$$y_0 = f(x_0).$$

Since $y_0$ is between zero and one, we may write

$$y_0 = 0.t_1t_2t_3 \cdot \cdot \cdot .$$

The values of $x$ may be written

$$x = \frac{x_1}{10} + \frac{x_3}{10^3} + \frac{x_5}{10^5} + \cdot \cdot \cdot + \frac{x_{2n-1}}{10^{2n-1}} + \cdot \cdot \cdot \qquad (5)$$

$$+ \frac{x_2}{10^2} + \frac{x_4}{10^4} + \frac{x_6}{10^6} + \cdot \cdot \cdot + \frac{x_{2n}}{10^{2n}} + \cdot \cdot \cdot .$$

---

[1] See Lebesgue, *Leçons sur l'intégration*, p. 90.

If $f(x)$ is to be different from zero the numerators in the first row must, from some point on, become periodic. Let $x_{2n-1}$ be the last numerator before the periodicity begins. Suppose a point $x'$ of $(a, b)$ is given by taking the first $2n-1$ terms of the fraction; that is, let

$$x' = \frac{x_1}{10} + \frac{x_2}{10^2} + \cdot \quad \cdot \quad \cdot + \frac{x_{2n-1}}{10^{2n-1}}.$$

Whatever digits be chosen as numerators following $x_{2n-1}$, the value of $x$ can differ from $x'$ by at most $\frac{1}{10^{2n-1}}$. By the proper choice of $n$ we can make

$$\frac{1}{10^{2n-1}} < | b - x' | ,$$

and hence the value of $x$ lies within $(a, b)$. In (5) we shall now replace those numerators following $x_{2n-1}$ having even subscripts by the digits used in defining $y_0$; that is, we shall define $x_0$ as follows:

$$x_0 = 0.x_1x_2 \cdot \quad \cdot \quad \cdot \, x_{2n-1}\, t_1x_{2n+1}\, t_2x_{2n+3}\, t_3 \cdot \quad \cdot \quad \cdot \,.$$

The point $x_0$ is then a point of the subinterval $(a, b)$ and $f(x_0)$ is the value $y_0$. But $y_0$ is any point on the $Y$-axis between zero and one, and hence $f(x)$ must take all values between zero and one as $x$ is given the values of the arbitrarily small subinterval $(a, b)$. By Theorem V, $f(x)$ must then be totally discontinuous in the interval $(0, 1)$.

This example gives an illustration of the type of functions, mentioned in Art. 26, which may take all intermediate values between any two of its values and still be totally discontinuous. For, since $f(x)$ takes all values between zero and one and can have no values greater than one, it must take in every subinterval all intermediate values between its values at the extremities of that subinterval.

THEOREM VI. *The points of a given interval* $(a, b)$ *at which* $f(x)$ *has ordinary discontinuities form at most an enumerable set.*

Denote by $E_\eta$ the set of points in $(a, b)$ at which $f(x)$ has a discontinuity of the first kind such that the saltus is greater than $\eta$. We shall first show that $E_\eta$ is an isolated set and hence by the corollary to Theorem IV, Art. 15 is enumerable. If $E_\eta$ is a finite set, it follows at once that it is enumerable. If $E_\eta$ is an infinite set, it must then have at least one limiting point. Let $x_0$ be such a point. In every neighborhood of $x_0$ there must be an infinite number of pairs of points dense at $x_0$ such that the difference of the functional values is greater than $\eta$. Moreover, these pairs of points can be so chosen as to lie on the same side of $x_0$. For example, suppose these points lie to the right of $x_0$; hence, the right-hand limit $\underset{x \to x_0+}{L} f(x)$ can not exist. It follows from the definition of an ordinary discontinuity that $x_0$ can not be a point of discontinuity of this type and hence can not be a point of $E_\eta$. But $x_0$ was any limiting point of $E_\eta$, and consequently $E_\eta$ can not contain its limiting points and must, therefore, be an isolated set. It is therefore an enumerable set.

Assign to $\eta$ a sequence of positive values

$$\eta_1, \eta_2, \cdot \cdot \cdot, \eta_n, \cdot \cdot \cdot,$$

where

$$\underset{n \to \infty}{L} \eta_n = 0.$$

The corresponding set of points $E_{\eta_n}$ is enumerable for each value of $n$. The totality of points of ordinary discontinuity of $f(x)$ in $(a, b)$ is then the sum of the sequence of enumerable point sets

$$E_{\eta_1}, E_{\eta_2}, \cdot \cdot \cdot, E_{\eta_n}, \cdot \cdot \cdot,$$

and hence by Theorem I of Art. 15 must be enumerable.

If we have given a function $f(x, y)$ of two variables which is continuous in each variable separately, it is possible for it to have a discontinuity with respect to the two variables taken together. For example, we have seen that the function

$$\begin{aligned} f(x, y) &= \frac{xy}{x^2 + y^2}, \ x \neq 0, \ y \neq 0, \\ &= 0, \ x = 0, \ y = 0 \end{aligned}$$

is continuous in $x$ and in $y$ but has an $(x, y)$-discontinuity at the origin. It is continuous in the two variables elsewhere.

We shall now consider the question as to whether such a function can be totally discontinuous in $x$ and $y$ together and still be continuous in each variable separately. As an auxiliary theorem in answering this question we have the generalization of Theorem II which for the case of two variables may be stated as follows.

THEOREM VII. *If $f(x, y)$ is continuous in each variable separately the points at which it can have a discontinuity in the two variables together, where the saltus is equal to or greater than a given number $A$, form a closed set.*

The proof differs from that of Theorem II only in that one makes use of the generalized form for two variables of Theorem IV, Art. 27, and it can be readily supplied by the reader. By aid of this theorem we can now establish the following theorem.

THEOREM VIII. *If $f(x, y)$ is continuous in each variable separately, the set of points at which $f(x, y)$ has an $(xy)$-discontinuity where the saltus is equal to or greater than an arbitrarily chosen number $A$ can be in no subregion everywhere dense.*

Let $E$ be the set of points where the $(xy)$-saltus of $f(x, y)$ is equal to or greater than $A$. We shall prove the theorem by assuming that $E$ is everywhere dense and showing that this assumption leads to a contradiction of the hypothesis that $f(x, y)$ is continuous in $x$ and in $y$. For example, let the points of $E$ be everywhere dense in the region bounded by the lines

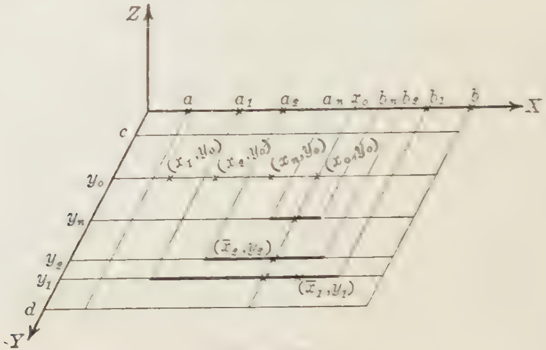


FIG. 27.

$$x = a,\ x = b,\ y = c,\ y = d.$$

From Theorem VII it follows that every point of this region is one where the saltus function $\omega(f, x, y)$ is equal to or exceeds $A$. On any line

$$y = y_0$$

lying in this region every point is such a point. Let $(x_1, y_0)$ be any point on $y = y_0$ such that $a < x_1 < b$. Since

$$\omega(f, x_1, y_0) \geqq A,$$

we can find some point $(\bar{x}_1, y_1)$ in the neighborhood of $(x_1, y_0)$ for which

$$| f(x_1, y_0) - f(\bar{x}_1, y_1) | > \frac{A}{2}. \tag{6}$$

Since $f(x, y_1)$ is continuous in $x$, there exists upon the line $y = y_1$, an interval $(a_1, b_1)$ lying within $(a, b)$ such that for every $x$ within $(a_1, b_1)$ we have

$$| f(\bar{x}_1, y_1) - f(x, y_1) | < \epsilon. \tag{7}$$

The value of $\epsilon$ is arbitrary and by taking it sufficiently small we have for every $x$ in $(a_1, b_1)$, upon combining (6) and (7),

$$| f(x_1, y_0) - f(x, y_1) | > \frac{A}{4}. \tag{8}$$

Within the projection of $(a_1, b_1)$ upon $y = y_0$ select any point $(x_2, y_0)$. This point is also one for which

$$\omega(f, x_2, y_0) \geqq A.$$

In any neighborhood of $(x_2, y_0)$ there must then exist a point $(\bar{x}_2, y_2)$ for which we have

$$| f(x_2, y_0) - f(\bar{x}_2, y_2) | > \frac{A}{2}. \tag{9}$$

As $(\bar{x}_2, y_2)$ may lie arbitrarily close to $y_0$, we may so select $y_2$ that it lies between $y_0$ and $y_1$. There must exist upon $y = y_2$ an interval $(a_2, b_2)$ lying within $(a_1, b_1)$ such that for every $x$ in this interval, we have

$$| f(\bar{x}_2, y_2) - f(x, y_2) | < \epsilon. \tag{10}$$

By the proper selection of $\epsilon$, we have by combining (9) and (10)

$$| f(x_2, y_0) - f(x, y_2) | > \frac{A}{4},$$

which holds for all values of $x$ in $(a_2, b_2)$. Select a point $(x_3, y_0)$ lying within $(a_2, b_2)$ and proceed as before. In this manner we obtain the general relation

$$| f(x_n, y_0) - f(x, y_n) | > \frac{A}{4}, \quad n = 1, 2, 3, \cdot \cdot \cdot, \tag{11}$$

for all values of $x$ within an interval $(a_n, b_n)$ and $y_n$ between $y_{n-1}$ and $y_0$. We have so selected the intervals $(a_n, b_n)$ that

$$a < a_1 < a_2 < \cdot \cdot \cdot < a_n < \cdot \cdot \cdot,$$
$$b > b_1 > b_2 > \cdot \cdot \cdot > b_n > \cdot \cdot \cdot .$$

We now impose upon the $a_n$ and $b_n$ the further condition that

$$\underset{n \to \infty}{L} a_n = \underset{n \to \infty}{L} b_n = x_0.$$

The point $x_0$ is common to all of the intervals $(a_n, b_n)$, and hence the points $x_1, x_2, x_3, \cdot \cdot \cdot$ are dense at $x_0$. Because of the continuity of $f(x, y_0)$ we have

$$\underset{x_n \to x_0}{L} f(x_n, y_0) = f(x_0, y_0).$$

Written as an inequality, this relation becomes for all values of $n$ greater than a certain number $m$

$$| f(x_0, y_0) - f(x_n, y_0) | < \epsilon. \tag{12}$$

As $x_0$ is common to all the intervals $(a_n, b_n)$ we have from (11)

$$| f(x_n, y_0) - f(x_0, y_n) | > \frac{A}{4}. \tag{13}$$

By combining the inequalities (12) and (13) we have for a sufficiently small $\epsilon$

$$|f(x_0, y_0) - f(x_0, y_n)| > \frac{A}{8}, \ n > m.$$

The values of the set $\{y_n\}$ were so chosen that they are dense at $y_0$. If the limit $\underset{y_n \to y_0}{L} f(x_0, y_n)$ exists we must have from the above inequality

$$\underset{y_n \to y_0}{L} f(x_0, y_n) = B \neq f(x_0, y_0).$$

Hence, either the limit $\underset{y \to y_0}{L} f(x_0, y)$ does not exist, or if it does it is different from the value of $f(x, y)$ at $(x_0, y_0)$. In either case $f(x, y)$ is not continuous in $y$ at $(x_0, y_0)$, which is in contradiction to the hypothesis that $f(x, y)$ is continuous in $x$ and in $y$. From this contradiction, the theorem follows.

THEOREM IX. *If $f(x, y)$ is continuous with respect to each variable separately in a closed region $R$, then it is at most point-wise discontinuous with respect to $(x, y)$ together, and the points of continuity in $(x, y)$ form a set having the cardinal number of the continuum.*

Suppose we have given a sequence of numbers $\{\eta_n\}$ such that

$$\eta_1 > \eta_2 > \cdot \cdot \cdot > \eta_n > \eta_{n+1} > \cdot \cdot \cdot ,$$

and

$$\underset{n \to \infty}{L} \eta_n = 0.$$

From Theorem VIII it follows that if $f(x, y)$ is continuous in $x$ and in $y$, the set of points $E_1$ where the saltus $\omega(f, x, y)$ is equal to or greater than $\eta_1$ can not be everywhere dense in $R$. Likewise, the set $E_2$ where $\omega(f, x, y)$ is equal to or greater than $\eta_2$ can not be everywhere dense in $R$, etc. In general the set $E_n$ where the saltus is equal to or greater than $\eta_n$ can not be everywhere dense whatever value is given to $n$. The set $E$ formed from the sum of these sets, that is

$$E = E_1 + E_2 + E_3 + \cdot \cdot \cdot + E_n + \cdot \cdot \cdot ,$$

is of the first category. This set $E$ is also the set of all the points of $(x, y)$-discontinuity of $f(x, y)$. Its complementary set $C(E)$ is the set of points where $f(x, y)$ is continuous in the two variables together. Since $E$ is of the first category, it follows from Theorem III of Art. 18, stated for two-dimensional sets, that $C(E)$ must be everywhere dense in $R$, and hence $f(x, y)$ is at most point-wise discontinuous in this region.

From Theorem V of Art. 18, it follows that the point set $C(E)$, that is the points where $f(x, y)$ is continuous in $x$ and $y$ together, also has the cardinal number of the continuum.

As we have already seen, a function $f(x, y)$ which is continuous in $x$ and in $y$ may still be discontinuous in $x$ and $y$ together. It will be shown in a later article[1] that these points of discontinuity may be everywhere dense in the given region. However, the points at which the $(x, y)$-saltus exceeds any given number can not be everywhere dense in any portion of that region.

In the previous article, we discussed the conditions that the limiting function $f(x, y_0)$ should be continuous in $x$, if $f(x, y)$ is continuous in $x$ for $y \neq y_0$ and $f(x, y_0)$ is defined by the limiting value

$$f(x, y_0) = \underset{y \to y_0}{L} f(x, y).$$

The limiting function $f(x, y_0)$ may be discontinuous in $x$ at certain points as the following exercise shows.

**Ex. 2.** Determine the points of discontinuity of $f(x, y_0)$ in the interval $(0, 1)$, where $y_0 = 0$ and where, for $x \neq 0$, $y \neq 0$, we have

$$f(x, y) = \frac{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} - 1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} + 1}.$$

We have $$f(x, 0) = \underset{y \to 0}{L} \frac{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} - 1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} + 1}, \quad 0 < x \leqq 1.$$

This limit has the value zero for

$$x = 1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{n}, \cdots,$$

since for these values $\sin\frac{\pi}{x}$ is zero. However, for $\frac{1}{2n+1} < x < \frac{1}{2n}$, we have $0 < \sin\frac{\pi}{x} \leqq 1$ and we get

$$f(x, 0) = \underset{y \to 0}{L} \frac{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} - 1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} + 1} = \underset{y \to 0}{L} \frac{1 - \dfrac{1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}}}}{1 + \dfrac{1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}}}} = 1.$$

On the other hand, for $\frac{1}{2n} < x < \frac{1}{2n-1}$, we have $0 > \sin\frac{\pi}{x} \geqq -1$ and hence

$$f(x, 0) = \underset{y \to 0}{L} \frac{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} - 1}{\left(1 + \sin\frac{\pi}{x}\right)^{\frac{1}{y}} + 1} = -1.$$

---

[1] See Art. 64.

Consequently the function $f(x, 0)$ is discontinuous at the points

$$x = 1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{n}, \cdots.$$

We shall now examine the limiting function $f(x, y_0)$ for points of discontinuity. Can this function be a totally discontinuous function of $x$? Are there points on the boundary $y = y_0$ at which $f(x, y)$ is continuous in the two variables $(x, y)$ together? If such points exist, what are some of the characteristic properties of the set of such points? The importance of these questions becomes apparent when we reflect that any property which the limiting function $f(x, y_0)$ may have holds for the special case where we consider the limiting function of a sequence of functions $\{f_n(x)\}$, or a function defined by an infinite series. We shall answer these questions by establishing the following theorems.

THEOREM X. *Let* $f(x, y)$ *be continuous in* $x$ *for* $y \neq y_0$ *and* $a \leqq x \leqq b$, *and suppose* $f(x, y_0)$ *to be defined by the relation*

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0). \qquad (a)$$

*Then* $f(x, y_0)$ *can be at most a point-wise discontinuous function of* $x$, *and hence the points of continuity form a set everywhere dense in* $(a, b)$ *and this set has the cardinal number of the continuum.*

Let $E$ be the set of points of discontinuity of $f(x, y_0)$ where the $x$-saltus is equal to or greater than any arbitrarily chosen, positive number $A$. We shall first show that the points of $E$ can not be everywhere dense in $(a, b)$. To prove this we shall assume the contrary to be true and prove that this assumption leads to a contradiction. Denote by $\omega(f, x, y_0)$ the saltus of $f(x, y_0)$ as a function of $x$. Since $E$ was assumed to be everywhere dense in $(a, b)$ then every point of this interval is a point of $E$.

Consider a sequence of points $\{x_n\}$ on the line $y = y_0$ satisfying the condition

$$| f(x_n, y_0) - f(x_{n+1}, y_0) | \geqq \frac{A}{2}. \qquad (14)$$

For each $x_n$ of this set the relation $(a)$ holds, and hence we have

$$| f(x_n, y_n) - f(x_n, y_0) | < \frac{\epsilon}{2}, \; y_0 < y_n < y_0 + \delta, \qquad (15)$$

where $n = 1, 2, 3, \cdots$, and the points $y_n$ are so chosen that the sequence $\{y_n\}$ is dense at $y_0$. Since $f(x, y)$ is continuous in $x$ for $y \neq y_0$, we have on each $y = y_n$ an interval $(\alpha_n, \beta_n)$, $n = 1, 2, 3, \cdots$ such that for all values of $x$ in this interval we have

$$|f(x, y_n) - f(x_n, y_n)| < \frac{\epsilon}{2}, \ \alpha_n < x < \beta_n. \tag{16}$$

Combining the inequalities (15) and (16) we obtain

$$|f(x, y_n) - f(x_n, y_0)| < \epsilon, \tag{17}$$

which holds for $y_n$ sufficiently close to $y_0$ and for all values of $x$ in $(\alpha_n, \beta_n)$, $n = 1, 2, 3, \cdots$. The intervals $(\alpha_n, \beta_n)$ may vary with $y_n$, but they can be so chosen that each lies within the preceding one, and the points $x_n$ may be so chosen that $x_n$ lies within the interval $(\alpha_n, \beta_n)$ as indicated in the accompanying figure.

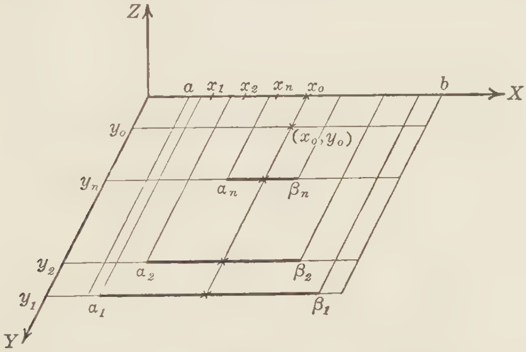


FIG. 28

As (17) holds for all integral values of $n$, we have also the inequality

$$|f(x_{n+1}, y_0) - f(x, y_{n+1})| < \epsilon, \tag{18}$$

where $y_0 < y_{n+1} < y_n$, and $x_n < x_{n+1} < \beta_n$. Combining the inequalities (14), (17) and (18), we obtain

$$|f(x, y_n) - f(x, y_{n+1})| > \frac{A}{4}, \tag{19}$$

provided $\epsilon$ is chosen less than $\frac{A}{8}$. Since each interval $(\alpha_n, \beta_n)$ lies within the preceding one, it follows that $\{x_n\}$ has at least one limiting point $x_0$ which can replace the $x$ in (19). Hence, we have for all values of $n$

$$|f(x_0, y_n) - f(x_0, y_{n+1})| > \frac{A}{4},$$

and consequently the limit given in ($a$) can not exist for $x = x_0$. From this contradiction, it follows that the points where $f(x, y_0)$ has a saltus function $\omega(f, x, y_0)$ equal to or greater than $A$ can not be everywhere dense in $(a, b)$; that is, the points of $E$ are not everywhere dense in this interval.

Consider now a sequence of values $\{\eta_n\}$, where

$$\eta_1 > \eta_2 > \cdot\ \cdot\ \cdot > \eta_n > \cdot\ \cdot\ \cdot,$$

and $\underset{n \to \infty}{L}\ \eta_n = 0$. Let $G$ denote the set of all points of discontinuity of $f(x, y_0)$. Denote by $E_1$ those points of $G$ where

$$\omega(f, x, y_0) \geqq \eta_1.$$

From the foregoing discussion, $E_1$ can not be a set of points everywhere dense on $y = y_0$. Similarly, let $E_2$ be those points of $G$ where the saltus is equal to or greater than $\eta_2$, etc. In general let $E_n$ be the set composed of those points of $G$ where the value of $\omega(f, x, y_0)$ is equal to or greater than $\eta_n$. We have then

$$G = E_1 + E_2 + \cdot\ \cdot\ \cdot + E_n\ \cdot\ \cdot\ \cdot,$$

where none of the sets $E_n$ are everywhere dense in the given interval. The set $G$ is then of the first category. The complementary set $C(G)$, that is the set of points where $f(x, y_0)$ is continuous in $x$, must be a set of the second category. It is then everywhere dense and moreover has the cardinal number of the continuum, as the theorem requires.

Let us consider the possibility of the existence of points on the line $y = y_0$ where the limiting function $f(x, y_0)$, as already defined, shall satisfy the condition

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L}\ f(x, y) = f(x_0, y_0).$$

Such a point may be conveniently characterized by saying that $f(x, y)$ converges uniformly[1] at $(x_0, y_0)$. In this connection we have the following theorem.

THEOREM XI. *If $f(x, y)$ is continuous in $x$ for $y \neq y_0$, $a \leqq x \leqq b$, and if $f(x, y_0)$ is defined by the limit $\underset{y \to y_0}{L}\ f(x, y)$, then the points on $y = y_0$ where $f(x, y)$ converges uniformly form a set which is everywhere dense in $(a, b)$ and has the cardinal number of the continuum.*

Let $(\alpha, \beta)$ be any subinterval of $(a, b)$. From Theorem X it follows that in some portion of $(\alpha, \beta)$, say $(\alpha_1, \beta_1)$ the $x$-saltus of

---

[1] See *Encyclop. der Math. Wiss.*, Vol. 2, C. 9, p. 1141.

$f(x, y_0)$ must be less than an arbitrarily chosen positive number $\epsilon$ at every point. We shall assume that in this interval $(\alpha_1, \beta_1)$ there exists no point at which we have $\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x_0, y) = f(x, y_0)$, and show that this assumption leads to a false conclusion. Then, at every point the $(x, y)$-saltus is different from zero and hence in some portion of this interval, say in $(\alpha_2, \beta_2)$, this saltus must exceed some positive number $A$. It then follows that in every neighborhood of any point $(x_2, y_0)$ of $(\alpha_2, \beta_2)$ there exists another point $(\bar{x}_2, y_2)$ such that

$$|f(\bar{x}_2, y_2) - f(x_2, y_0)| > \frac{A}{2}, \; y_0 < y_2 < y_0 + \delta.$$

Because of the continuity of $f(x, y_2)$ in $x$, we have an interval $(\alpha_3, \beta_3)$ on $y = y_0$ and lying within $(\alpha_2, \beta_2)$ such that for all values of $x$ in $(\alpha_3, \beta_3)$ we have

$$|f(x, y_2) - f(\bar{x}_2, y_2)| < \frac{\epsilon}{2}.$$

Let $x_3$ be any point in $(\alpha_3, \beta_3)$. In every neighborhood of $(x_3, y_0)$ there exists a point $(\bar{x}_3, y_3)$ such that

$$|f(\bar{x}_3, y_3) - f(x_3, y_0)| > \frac{A}{2}, \; y_0 < y_3 < y_2.$$

Again, because of the continuity of $f(x, y_3)$ in $x$, there exists within $(\alpha_3, \beta_3)$ an interval $(\alpha_4, \beta_4)$ such that for all values $x$ in this interval, we have

$$|f(x, y_3) - f(\bar{x}_3, y_3)| < \frac{\epsilon}{2}.$$

Continuing in this manner, we obtain a sequence of intervals $(\alpha_n, \beta_n)$, each lying within the preceding, such that

$$|f(\bar{x}_n, y_n) - f(x_n, y_0| > \frac{A}{2}, \tag{20}$$

$$|f(x, y_n) - f(\bar{x}_n, y_n| < \frac{\epsilon}{2}, \tag{21}$$

where the set of values $\{y_n\}$ is dense at $y_0$ and the values of the set $\{x_n\}$ may be so chosen as to be dense at $x_0$. Then, since at $(x_0, y_0)$ the $x$-saltus of $f(x, y_0)$ is less than $\epsilon$, we have an interval $(x_0 - \delta_{y_0}, x_0 + \delta_{y_0})$ such that for each $x_n$ in this interval we have

$$|f(x_n, y_0) - f(x_0, y_0)| < \frac{\epsilon}{2}. \tag{22}$$

Combining (20), (21) and (22) we have for $\epsilon < \frac{A}{4}$

$$|f(x, y_n) - f(x_0, y_0)| > \frac{A}{4},$$

where $x$ is taken in a sufficiently small neighborhood of $x_0$. It then must hold for $x_0$ itself; that is, we have

$$|f(x_0, y_n) - f(x_0, y_0)| > \frac{A}{4},$$

for each $y_n$ of the set $\{y_n\}$ sufficiently close to $y_0$. It follows then that the limit $\underset{y \to y_0}{L} f(x_0, y)$ can not exist, which is contrary to the given hypothesis. Hence, in the interval $(\alpha, \beta)$ there must exist at least one point where

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f(x, y) = f(x_0, y_0);$$

that is, there must exist at least one point where $f(x, y)$ converges uniformly. As $(\alpha, \beta)$ is any subinterval of $(a, b)$, it follows that these points are everywhere dense in $(a, b)$.

It may be shown in the same manner as in the proof of Theorem X that these points must have the cardinal number of the continuum. This part of the proof is left to the reader.

In the foregoing discussion one may replace $y$ in $f(x, y)$ by a set of constants $\{y_n\}$ dense at $y_0$. We have then a sequence of functions of $x$ alone $\{f(x, y_n)\}$ which can be conveniently written

$$\phi_1(x),\ \phi_2(x),\ \cdots,\ \phi_n(x),\ \cdots.$$

All of these functions $\phi_n(x)$ are continuous in $x$, and the limit of the sequence is $\phi(x)$, where

$$\phi(x) \equiv f(x, y_0),$$

which at most can be point-wise discontinuous. It follows from the properties of the limiting function that the limit of a sequence of continuous functions is at most point-wise discontinuous. The converse is also true, namely that every point-wise discontinuous function may be represented as the limit of a sequence of continuous functions. Hence, a necessary and sufficient condition that a function can be represented as the limit of a sequence of continuous functions is that it be at most point-wise discontinuous.[1] It follows also that every function which is the limit of a sequence of polynomials is necessarily continuous.

The consideration of the character of a function which may be represented as the limit of a sequence of functions leads to what is known as the **Baire classification of functions.**[2] In class zero he puts all continuous functions. Class one is composed of all those functions which are not continuous but are the limits of sequences of continuous functions. In class two are put those functions which are the limits of sequences of functions of class one and do not belong to class zero or class one. In general, a function is said to be of class $\alpha$, if it is the limit of a sequence of functions of class $\alpha - 1$ and does not belong to any one of the classes $0, 1, 2, \cdots, \alpha - 1$.

---

[1] See Baire, *Leçons sur les fonctions discontinues*, p. 124.

[2] See Baire, *ibid.*, p. 126.

Any point-wise discontinuous function is an illustration of a function of class one. It is of interest to note that it has been shown that there exist functions of every Baire class of finite order,[1] and that functions may be defined which do not belong to any such class.[2]

THEOREM XII. *A bounded function of class* $\alpha > 0$ *may be expressed as the limit of a sequence of functions of class less than* $\alpha$ *which have the same bounds as the given functions.*

By hypothesis the given function $f(x)$ is bounded. Denote the least upper bound by $A$ and the greatest lower bound by $B$. If the functions of the sequence $\{f_n(x)\}$ defining $f(x)$ are such that their bounds do not exceed those of $f(x)$ for finite values of $n$, they are, nevertheless, equal to those of $f(x)$ in the limit as $n$ becomes infinite. In case the bounds of $f_n(x)$ do exceed those of $f(x)$ for some values of $n$, we may replace $f_n(x)$ by another function $\phi_n(x)$ defined as equal to $f_n(x)$ for those values of $x$ for which $A \geqq f(x) \geqq B$ and equal to $A$ for those values of $x$ for which $f_n(x) > A$ and likewise equal to $B$ for those values of $x$ for which $f_n(x) < B$. Having replaced $f_n(x)$ by $\phi_n(x)$, we can in this case also conclude that $f(x)$ may be defined as the limit of a sequence of functions having the same least upper and greatest lower bounds as $f(x)$ itself; that is, by the sequence $\{\phi_n(x)\}$ which has $A$ and $B$ as its least upper and greatest lower bounds, respectively.

It is to be noted that the class of the function $f(x)$ is not increased by thus restricting the bounds of the functions of the sequence defining it. For example, it will be seen at once that this is true where $f(x)$ is a function of Baire's class zero, since in this case $f(x)$ is continuous and is the limit of continuous functions. That the same is true of functions of any class follows at once from the fact that by thus restricting $f_n(x)$ we do not introduce any new points of discontinuity and may indeed reduce the number of such points. It follows then that $f(x)$ is of class $\alpha$ or less irrespective of whether it is defined by the sequence $\{f_n(x)\}$ or by the restricted sequence $\{\phi_n(x)\}$. With this conclusion the proof is completed.

---

[1] See Borel, *Leçons sur les fonctions de variables reelles,* p. 156; also Lebesgue, *Liouville's Journ.* (6) Vol. 1 (1905) p. 139.

[2] See Lebesgue, *Liouville's Journ.* (6), Vol. 1 (1905), p. 213; also Sierpinski, *Fundamenta Mat.*, Vol. V (1924), p. 87.

THEOREM XIII. *Any function of class* $\alpha > 0$ *is the limit of a sequence of bounded functions of class less than* $\alpha$.

If the given function is bounded, then the theorem is identical with the preceding. If the given function $f(x)$ is not bounded, the proof is similar to that of the preceding theorem except that we replace $f_n(x)$ by $\phi_n(x)$ defined to be equal to $f_n(x)$ where $-n \leqq f_n(x) \leqq n$ and equal to $-n$ where $f_n(x) < -n$ and to $n$ where $f_n(x) > n$.

THEOREM XIV. *The sum or the product of two functions which are finite and of class* $\alpha$ *is a function of class not greater than* $\alpha$.

The theorem may, of course, be extended to any finite number of functions. It follows at once in case the given functions $f(x)$ and $F(x)$ are continuous; that is, of Baire's class zero, since the sum or product of two continuous functions is a continuous function.

Let us assume $f(x)$ and $F(x)$ to be finite and of class $\alpha > 0$. We then have

$$f(x) = \underset{n \to \infty}{L} f_n(x),\ F(x) = \underset{n \to \infty}{L} F_n(x), \tag{23}$$

where $f_n(x)$, $F_n(x)$ are of class less than $\alpha$. We shall now assume that the theorem holds for classes less than $\alpha$ and show that it holds for class $\alpha$. The theorem then follows, since it holds for continuous functions, that is for class zero, and hence for class one, class two, etc. Since it holds for classes less than $\alpha$, it follows that $f_n(x) + F_n(x)$ is of class less than $\alpha$. By adding the relations given in (23), we have

$$f(x) + F(x) = \underset{n \to \infty}{L} \{f_n(x) + F_n(x)\};$$

that is, the sum $f(x) + F(x)$ is the limit of a sequence of functions of class less than $\alpha$ and hence is of class not greater than $\alpha$, which was to be shown. In case the singularities of $f_n(x)$ and $F_n(x)$ are such that by adding these two functions some of them disappear, it may be possible that the class of the sum will be less than that of either function and consequently the class of $f(x) + F(x)$ may be less than $\alpha$.

By multiplying the two equations given in (23) member by member, we can by a similar argument establish the theorem for the product of two functions.

## EXERCISES

**1.** Show that $f(x) = \frac{1}{x}\sin\frac{1}{x}$, for $x \neq 0$, where $f(0) = 0$, is finite for every value of $x$ in the interval $(-1, 1)$ but is not bounded. Plot the graph. Determine the point of discontinuity. What is the amount of saltus?

**2.** Evaluate the following limits, or show that the limits do not exist.

(a) $\underset{x\to\infty}{L}\, \frac{1}{2x}\cos\frac{\pi x}{2}$; (b) $\underset{x\to\infty}{L}\, \frac{x^{2x}}{(1+x^2)^x}$; (c) $\underset{\substack{x\to 0\\ y\to 0}}{L}\, (x+y)\frac{y+(x+y)^2}{y-(x+y)^2}$;

(d) $\underset{n\to\infty}{L}\, nx^2e^{-nx^2}$; (e) $\underset{\substack{x\to 0\\ y\to 0}}{L}\, \log\frac{a(1-e^x)}{a-y^x}$.

**3.** Given the function $f(x) = \underset{m\to\infty}{L}\left[\underset{n\to\infty}{L}\,(\cos m!\,\pi x)^{2n}\right]$. Show that $f(x)$ is one for all rational values of $x$ and zero for all irrational values of $x$, thus being totally discontinuous in the interval $(0, 1)$. Is this function a semi-continuous function?

**4.** Let $f(x) = \frac{1}{1+e^{\frac{1}{x}}} + \sin\frac{1}{|x|}$, for $x \neq 0$, and let $f(x) = 0$, for $x = 0$. Discuss the discontinuity of $f(x)$ at the origin. What is the maximum value of $f(x)$ at $x = 0$?

**5.** Given $f(x) = \frac{1}{1 + e^{\frac{1}{\sin n!\pi x}}}$. Show that $f(x)$ can be made discontinuous at any rational point in the interval $(0, 1)$ by the proper selection of $n$. Is it point-wise discontinuous?

**6.** Show that if a function is continuous the number of oscillations greater than $A$ must be finite in any finite interval.

**7.** In the closed interval $(-1, 1)$ let $f(x)$ be defined as $x^2 \sin\frac{1}{x^2}$ for $x \neq 0$ and as zero for $x = 0$. In the given interval is this function bounded? Is it continuous? Is it uniformly continuous? Is it absolutely continuous? Is it semi-continuous? What is the maximum value of its saltus?

**8.** Given $f(x, y) = \frac{xy}{x^3 + y^2}$, when either or both of the variables are different from zero, and let $f(0, 0) = 0$. Show that in the region where $x \geqq 0$, $y \geqq 0$. $f(x, y)$ is bounded for each constant value of $x$ and for each constant value of $y$, but not bounded in $(x, y)$.

**9.** Given $f(x, y) = \log\,(x^2 + y^2)^{\frac{y^2}{2}}$, if either or both of the variables are different from zero. Moreover, let $f(0, 0) = 0$. Examine the function at the origin for continuity as to $x$, as to $y$, as to $(x, y)$ together.

**10.** Show that every rational function is continuous at every point where it is neither infinite nor indeterminate. What is the most general set of points of discontinuity which such a function may have?

**11.** Set up a function $f(x, y)$ such that when one puts $x = \rho\cos\theta$, $y = \rho\sin\theta$, a function $F(\rho, \theta)$ is obtained such that $\underset{\rho\to 0}{L}\, F(\rho, \theta) = 0$ for each value of $\theta$ and yet $f(x, y)$ is discontinuous in $(x, y)$ at the origin.

**12.** Give a function $f(x, y)$ which is continuous in $x$ and in $y$ but which does not converge uniformly to the limiting function $f(x, 0)$.

**13.** Is it possible to have a point-wise discontinuous function which has points of discontinuity at a set which is perfect and not everywhere dense? Give the reason for your conclusion.

**14.** Show that if $f(x)$ is continuous with respect to a bounded, perfect set $E$, then it is uniformly continuous with respect to $E$ (Compare Theorem I, Art. 26),

**15.** Given $f(x, y) = \frac{xy^2}{x^2 + y^4}$, $y \neq 0$ and all values of $x$, and let $f(0, 0) = 0$. Test $f(x, y)$ for quasi-uniform convergence to the limiting function $f(x, 0)$. What can one say concerning the continuity of the limiting function?

**16.** Prove Theorem I, Art. 26 by aid of the Heine-Borel Theorem.

**17.** Could one have a point-wise discontinuous function $f(x)$ whose points of continuity form a discrete set? an enumerable set? a set of the first species? a set of the first category? a closed set? a perfect set? a closed set of measure zero?

**18.** If $f(x)$ is a rational integral function defined for the interval $(a, b)$, does it take all values between $f(a)$ and $f(b)$? Is it uniformly continuous in $(a, b)$? Does it have a maximum and a minimum in $(a, b)$? Can it be represented as the limit of a sequence of continuous functions?

**19.** Given $f(x, y) = e^{-\frac{1}{x^2y^2}} \cos \frac{\pi}{xy}$ for y $\neq 0$, where we put $f(x, 0) = \underset{y \to 0}{L} f(x, y)$. Does $f(x, y)$ converge uniformly to $f(x, 0)$? Does it converge quasi-uniformly? Is the limiting function $f(x, 0)$ discontinuous? What is the Baire class to which $f(x, 0)$ belongs?

**20.** Given a function $f(x, y)$ which is continuous in $x$ and in $y$ in a closed region $R$. What statements can you make concerning the projection upon the $X$-axis and the $Y$-axis of its points of $(x, y)$-discontinuities?

# CHAPTER IV

## DERIVATIVES AND THEIR PROPERTIES

**29. Definition of a derivative and of a derivative number.** If we denote by $\Delta x$ an increment of $x$, then $f(x)$ is said to have a **finite derivative** at the point $x_0$ provided the limit

$$\underset{\Delta x \to 0}{L} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{1}$$

exists. The derivative is also said to exist if this limit becomes definitely infinite; that is, if as $\Delta x$ approaches zero the value of the above difference quotient increases indefinitely without oscillating between positive and negative values. In this case the function is said to have an **infinite derivative** at the given point. The derivative is often spoken of as the differential coefficient and is denoted by the symbols

$$D_x y, \frac{dy}{dx}, f'(x_0).$$

The aggregate of values of the derivative for all values of $x$ under consideration is called the **derived function** and may be denoted by $f'(x)$.

In order that $f(x)$ shall have a finite derivative, it is necessary that the numerator of (1) shall vanish as $\Delta x$ approaches zero; that is, $f(x)$ must be continuous at $x_0$. For example, the function

$$\begin{aligned} f(x) &= \sin \frac{1}{x} \quad x \neq 0, \\ &= 0, \qquad x = 0, \end{aligned}$$

can not have a finite derivative at the origin; for, although the function is finite in the neighborhood of the origin it is discontinuous at that point. It can not have an infinite derivative at the origin because the limit given in (1) is not definitely infinite.

If the difference quotient has a limit as $\Delta x$ approaches zero through positive values, we say that $f(x)$ has a **right-hand derivative** at $x_0$; if the limit exists as $\Delta x$ approaches zero through negative values, then $f(x)$ is said to have a **left-hand derivative** at $x_0$. Either or both the right-hand and left-hand derivatives may exist without the function having a derivative at the point. If the right-hand and left-hand derivatives are equal, then the common value of the two is the derivative at the point. In the case of infinite derivatives the right-hand and the left-hand derivatives must have the same algebraic sign.

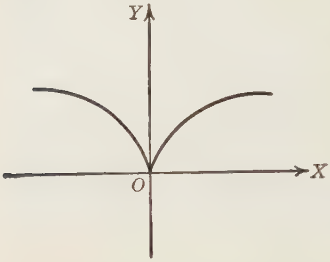


FIG. 29.

For example, if we have given

$$y = x^{\frac{2}{3}},$$

the corresponding curve has a cusp at the origin as shown in Fig. 29. Although the curve has a tangent at the origin, the derivative $\frac{dy}{dx}$ can not be said to exist at this point, since the right-hand derivative is plus infinity, while the left-hand derivative is negative infinity; for, we have for positive values of $\Delta x$

$$\underset{\Delta x \to 0}{L} \frac{f(0 + \Delta x) - f(0)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{(\Delta x)^{\frac{2}{3}}}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{1}{(\Delta x)^{\frac{1}{3}}} = \infty,$$

and on the other hand for negative values of $\Delta x$ we have

$$\underset{\Delta x \to 0}{L} \frac{f(0 - \Delta x) - f(0)}{- \Delta x} = \underset{\Delta x \to 0}{L} \frac{(- \Delta x)^{\frac{2}{3}}}{(- \Delta x)} = \underset{\Delta x \to 0}{L} \frac{-1}{(\Delta x)^{\frac{1}{3}}} = -\infty.$$

The derivative $\frac{dx}{dy}$ does exist, however, at the origin and is zero.

Suppose that we put for convenience

$$F(x_0, \Delta x) \equiv \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}, \ \Delta x \neq 0,$$

and consider the oscillation of $F(x_0, \Delta x)$ as $\Delta x$ takes all values subject to the condition that $x_0 + \Delta x$ shall lie in a given interval $(a, b)$, $x_0$ being a given point in that interval. We shall denote by $M(F, \Delta x, x_0)$ and $m(F, \Delta x, x_0)$ the least upper and the greatest lower bounds, respectively, of $F(x_0, \Delta x)$ as $\Delta x$ takes all values in the neighborhood of zero. Then, if $f'(x_0)$ exists, the saltus of $f'(x)$ at $x_0$ may be written

$$\omega(f', x_0) = M(F, x_0) - m(F, x_0), \tag{2}$$

where $M(F, x_0)$ and $m(F, x_0)$ are respectively the limits of $M(F, \Delta x, x_0)$ and $m(F, \Delta x, x_0)$ as $\Delta x$ approaches zero.

It is often convenient to distinguish between the right-hand and the left-hand limits of $M(F, \Delta x, x_0)$ and $m(F, \Delta x, x_0)$ as $\Delta x$ approaches zero. For that purpose put

$$D^+f(x_0) = \underset{\Delta x \to 0^+}{L} M(F, \Delta x, x_0), \tag{3}$$

$$D_+f(x_0) = \underset{\Delta x \to 0^+}{L} m(F, \Delta x, x_0).$$

The value $D^+f(x_0)$ is called an **upper right-hand derivative number** [1] of $f(x)$ at $x_0$, and $D_+f(x_0)$ is called the **lower right-hand derivative number** of $f(x)$ at $x_0$. Likewise, the relations

$$D^-f(x_0) = \underset{\Delta x \to 0^-}{L} M(F, \Delta x, x_0), \tag{4}$$

$$D_-f(x_0) = \underset{\Delta x \to 0^-}{L} m(F, \Delta x, x_0),$$

define respectively the **upper left-hand derivative number** and the **lower left-hand derivative number** of $f(x)$ at $x_0$. The four numbers, $D^+f(x_0)$, $D_+f(x_0)$, $D^-f(x_0)$, $D_-f(x_0)$, are also often spoken of as the four derivates at the point, or the upper and lower right-hand or left-hand derivatives. A derivative number may be said to be $+\infty$ or $-\infty$ if one of the foregoing limits becomes infinite with a definite sign as $\Delta x$ approaches zero through positive or negative values. One may call the aggregate of the values of the derivative numbers the upper and lower derivative functions.

If the upper and the lower derivative numbers are all equal at any point $x_0$, then of course the derivative $f'(x_0)$ exists and is the common value of the derivative numbers. In fact, the derivative at a point may be defined as this common value. All the rules for differentiation as given in elementary calculus can be readily developed upon the basis of such a definition.[2]

The geometric significance of the derivative numbers may be readily shown. In Fig. 30, let $P$ be any point on the curve $y = f(x)$, and let $x_0$ be the corresponding value of $x$. Suppose $Q$ to be any other point on the given curve and denote its abscissa by $x_0 + \Delta x$. For the right-hand derivative numbers, we shall consider $\Delta x$ as positive. Join $P$ with the point $Q$ of the given arc by a straight line. If $Q$ takes different positions on the curve, all lines $PQ$ will fall within some sector, say $APB$. As $\Delta x$ decreases and approaches zero, this sector never increases although it may decrease. Let $PA'$ and $PB'$ be the limiting positions of the lines $PA$ and $PB$, respectively, as $\Delta x$ approaches zero. The tangents of the angles which $PA'$ and $PB'$ make with the positive $X$-axis may be taken as the geometric interpretations of the derivative numbers $D^+f(x)$ and $D_+f(x)$. Corresponding interpretations of the left-hand derivatives also exist.

---

[1] The notion of upper and lower derivative numbers was first considered by DuBois-Reymond, *Abh. Akad. München*, Vol. 12 (1876), and Dini, *Fondamenti per la teorica delle funzioni di variabili reali* (1878). The notation used in the text is due to Scheeffer, *Acta. Math.* Vol. 5 (1884), p. 52.

[2] See Caratheodory, *Vorlesungen über Reelle Funktionen*, pp. 518–627.

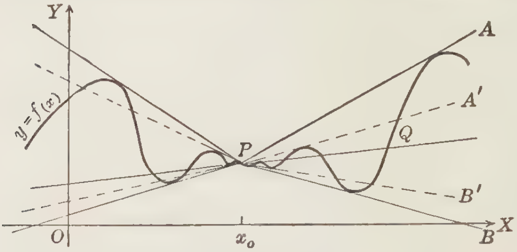


FIG. 30.

In discussing the properties of a function, the derivative numbers may often replace the ordinary derivative. For example, if the right-hand derivative number $D_+f(x)$ is positive for $x = x_0$, the function is increasing with $x$ at that point. If the left-hand derivative number $D_-f(x)$ is positive for $x = x_0$, then $f(x)$ is increasing as $x$ decreases. If we have

$$D^+f(x_0) < 0, \quad D_-f(x_0) > 0,$$

then $f(x)$ has a maximum value at $x_0$. If, on the other hand,

$$D^-f(x_0) < 0, \quad D_+f(x_0) > 0,$$

then $f(x)$ has a minimum value at $x_0$. If for all values of $x$ the two right-hand derivative numbers of a continuous function $f(x)$ are never both negative or both positive in the given interval, then in that interval $f(x)$ can neither increase nor decrease and hence must be a constant. In the discussions to follow we shall be interested only in the properties of derivative numbers of continuous functions.[1]

The derived function $f'(x)$ may in turn have a derivative, which is called the **second derivative** of $f(x)$ and is denoted symbolically by

$$D_x^2y, \frac{d^2y}{dx^2}, f''(x_0).$$

If $f'(x)$ exists at each point of an interval $(x_0 - \delta, x_0 + \delta)$ and is continuous at $x_0$, then the second derivative at $x = x_0$ is given by the limit

$$\underset{\Delta x \to 0}{L} \frac{f'(x_0 + \Delta x) - f'(x_0)}{\Delta x}, \tag{5}$$

---

[1] For a discussion of the properties of derivative numbers of discontinuous functions, see W. Sierpinski, *Fundamenta. Math.*, Vol. 3 (1922), pp. 123–127; also St. Banach, *ibid.*, pp. 128–132.

or by the iterated double limit

$$\underset{\Delta_2 x\to 0}{L}\ \underset{\Delta_1 x\to 0}{L}\ \frac{f(x_0+\Delta_1 x+\Delta_2 x)-f(x_0+\Delta_1 x)-f(x_0+\Delta_2 x)+f(x_0)}{\Delta_1 x\,\Delta_2 x}. \qquad (6)$$

One might infer from (6) that it would be permissible to put $\Delta_1 x$ equal to $\Delta_2 x$ and thus obtain $f''(x_0)$ from the limit

$$\underset{\Delta x\to 0}{L}\ \frac{f(x_0+2\Delta x)-2f(x_0+\Delta x)+f(x_0)}{(\Delta x)^2}. \qquad (7)$$

This method is not always possible, as the following illustration shows.

Ex. **1.** Given $f(x) = x^3 \cos\frac{1}{x}$, for $x \neq 0$, and let $f(0) = 0$. Consider the existence of $f''(0)$.

The first derivative is $f'(x) = 3x^2 \cos\frac{1}{x} + x \sin\frac{1}{x}$, for $x \neq 0$,

$$= 0, \text{ for } x = 0.$$

The derived function $f'(x)$ is continuous at $x = 0$ and exists at every point in the neighborhood of this point; hence, if $f''(0)$ exists, it is given by the limit

$$\underset{\Delta x\to 0}{L}\ \frac{f'(0+\Delta x)-f'(0)}{\Delta x} = \underset{\Delta x\to 0}{L}\left[3\Delta x\cos\frac{1}{\Delta x}+\sin\frac{1}{\Delta x}\right].$$

But this limit does not exist, since the limit $\underset{\Delta x\to 0}{L}\ \sin\frac{1}{\Delta x}$ does not exist. From (7) we have

$$\underset{\Delta x\to 0}{L}\ \frac{f(0+2\Delta x)-2f(0+\Delta x)+f(0)}{(\Delta x)^2} = \underset{\Delta x\to 0}{L}\left[8\Delta x\cos\frac{1}{2\Delta x}-2\Delta x\cos\frac{1}{\Delta x}\right] = 0,$$

and consequently $f''(0)$ is not given by this limit.

We may, however, state the following theorem.

Theorem[1] I. *Let $f(x)$ be a continuous function in the interval $(a, b)$. If the first and second derivatives $f'(x)$, $f''(x)$ exist and are finite for $x = x_0$, where $a < x_0 < b$, then it follows that*

$$f''(x_0) = \underset{\Delta x\to 0}{L}\ \frac{f(x_0+\Delta x)-2f(x_0)+f(x_0-\Delta x)}{(\Delta x)^2},$$

*or*

$$f''(x_0) = \underset{\Delta x\to 0}{L}\ \frac{f(x_0+2\Delta x)-2f(x_0+\Delta x)+f(x_0)}{(\Delta x)^2}.$$

We shall consider the auxiliary function

$$\phi(x) = f(x) - \frac{k}{2}x^2, \qquad (8)$$

where $k$ is an arbitrary constant. Since $f''(x)$ exists at $x_0$, it follows that $f'(x)$ must exist not only at $x_0$, but at every point in a neighborhood $(x_0 - \delta, x_0 + \delta)$ of $x_0$. If we choose $\Delta x$ sufficiently small so that $x_0 \pm \Delta x$ lies within this subinterval, we may write, by aid of the Law of the Mean,[1]

---

[1] See Dini, *Grundlagen*, Art. 174.

$$\phi(x_0 + \Delta x) - \phi(x_0) = \Delta x\, \phi'(x_0 + \theta_1 \Delta x),\ 0 < \theta_1 < 1,$$
$$\phi(x_0 - \Delta x) - \phi(x_0) = -\Delta x\, \phi'(x_0 - \theta_2 \Delta x),\ 0 < \theta_2 < 1.$$

Adding these equations, we have, after dividing by $(\Delta x)^2$,

$$\frac{\phi(x_0 + \Delta x) - 2\phi(x_0) + \phi(x_0 - \Delta x)}{(\Delta x)^2} = \frac{\phi'(x_0 + \theta_1 \Delta x)}{\Delta x} - \frac{\phi'(x_0 - \theta_2 \Delta x)}{\Delta x}$$
$$= \frac{\phi'(x_0 + \theta_1 \Delta x) - \phi'(x_0)}{\Delta x} + \frac{\phi'(x_0 - \theta_2 \Delta x) - \phi'(x_0)}{-\Delta x}$$
$$= \theta_1 \frac{\phi'(x_0 + \theta_1 \Delta x) - \phi'(x_0)}{\theta_1 \Delta x} + \theta_2 \frac{\phi'(x_0 - \theta_2 \Delta x) - \phi'(x_0)}{-\theta_2 \Delta x}. \qquad (9)$$

From (8) we have $\phi''(x_0) = f''(x_0) - k.$

As $k$ is an arbitrary constant, we will now assign to it the value $f''(x_0)$, and hence we have

$$\phi''(x_0) = 0.$$

In (9) the fractions in the second member define in the limit, as $\Delta x$ approaches zero, the right-hand and left-hand derivatives, respectively, of $\phi'(x)$ at $x = x_0$. But as $\phi''(x_0)$ is zero, both the right-hand and the left-hand derivatives must be equal to zero, and we have

$$\underset{\Delta x \to 0}{L} \frac{\phi(x_0 + \Delta x) - 2\phi(x_0) + \phi(x_0 - \Delta x)}{(\Delta x)^2} = 0.$$

Replacing $\phi(x)$ by its equivalent given in (8), we have

$$\underset{\Delta x \to 0}{L} \frac{f(x_0 + \Delta x) - 2f(x_0) + f(x_0 - \Delta x)}{(\Delta x)^2} = k = f''(x_0),$$

which is one of the conclusions stated in the theorem.

We may also write, by aid of the Law of the Mean,

$$\phi(x_0 + 2\Delta x) - \phi(x_0) = 2\Delta x \phi'(x_0 + 2\theta_1 \Delta x),\ 0 < \theta_1 < 1,$$
$$\phi(x_0 + \Delta x) - \phi(x_0) = \Delta x \phi'(x_0 + \theta_2 \Delta x),\ 0 < \theta_2 < 1.$$

By a process of reasoning similar to that used in the first part of this proof, we are led to the second conclusion of the theorem; namely,

$$\underset{\Delta x \to 0}{L} \frac{f(x_0 + 2\Delta x) - 2f(x_0 + \Delta x) + f(x_0)}{(\Delta x)^2} = f''(x_0).$$

---

[1] See Art. 31.

**30. Continuous functions having no derivative.** It was pointed out in the previous article that a function must be continuous at a given point in order to have a finite derivative at that point. The converse is not true, as the following example shows.

**Ex. 1.** Given the function $f(x) = x \sin \frac{1}{x}$, $x \neq 0$, where $f(0) = 0$. Show that $f(x)$ is continuous at $x = 0$, but has no derivative at that point.

As we have seen (Ex. 2, Art. 23)

$$\underset{x \to 0}{L}\ x \sin \frac{1}{x} = 0,$$

and hence $f(x)$ is continuous at this point, since $\underset{x \to 0}{L}\ f(x) = f(0)$. The function, however, does not have a derivative at this point; for,

$$\underset{\Delta x \to 0}{L} \frac{f(0 + \Delta x) - f(0)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{\Delta x \sin \frac{1}{\Delta x}}{\Delta x} = \underset{\Delta x \to 0}{L} \sin \frac{1}{\Delta x},$$

and this limit does not exist, since the sine of $\frac{1}{\Delta x}$ oscillates between the values plus one and minus one (see Fig. 11).

It must not be assumed that the derivative failed to exist in the foregoing illustration merely because of the infinite number of oscillations of $f(x)$ in the neighborhood of the origin. The following example shows that $f(x)$ may have an infinite number of oscillations in the neighborhood of a point and still have a derivative.

**Ex. 2.** Given $f(x) = x^2 \sin \frac{1}{x}$, $x \neq 0$, where $f(0) = 0$. Find the derivative of $f(x)$ for $x = 0$.

We have

$$\underset{x \to 0}{L}\ x^2 \sin \frac{1}{x} = 0 = f(0),$$

and hence $f(x)$ is continuous at the origin. The derivative exists at this point; for, we have

$$f'(0) = \underset{\Delta x \to 0}{L} \frac{f(0 + \Delta x) - f(0)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{(\Delta x)^2 \sin \frac{1}{\Delta x}}{\Delta x} = \underset{\Delta x \to 0}{L}\ \Delta x \sin \frac{1}{\Delta x} = 0.$$

Formerly, it was assumed that all continuous functions must have a derivative at every point, or at least with the exception of at most a finite number of points. The investigations of Riemann[1] and Bolzano[2] showed that functions exist which fail to have a derivative at certain sets of points. Weierstrass[3] was the first to give an example of a function which is continuous throughout an interval and still has no derivative at any point of that interval.

---

[1] *Göttingen Dissertation* (1854); Werke, 2d Ed., p. 229.

[2] See, *Encyklopädie der Math. Wiss.* Vol. II, p. 1091.

[3] Announced in a lecture course at Berlin in 1861. See, *Abh. aus der Funktionenlehre*, p. 97. This function is defined by means of an infinite series and the conditions stated by Weierstrass have been extended since by various writers. For further literature on this subject, see, *Encyklopädie der Math. Wiss.* Vol. II, p. 1091, footnote 736.

We shall now show geometrically how such a function can be set up; that is, we shall show how one may construct the successive approximations to a curve which is continuous in a given interval but which has no definite direction at any point.[1]

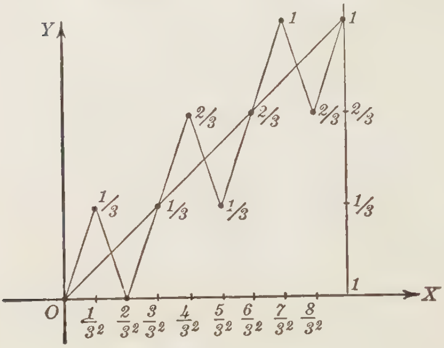

Y
X
O
1/3
2/3
1
1/3
2/3
1
2/3
1/3
2/3
1/3
1
$\frac{1}{3^2}$ $\frac{2}{3^2}$ $\frac{3}{3^2}$ $\frac{4}{3^2}$ $\frac{5}{3^2}$ $\frac{6}{3^2}$ $\frac{7}{3^2}$ $\frac{8}{3^2}$

FIG. 31.

Upon the $X$-axis lay off from the origin the interval (0, 1) and upon the perpendicular through the point $x = 1$ lay off a unit distance above the $X$-axis. Divide the unit interval thus formed on the perpendicular into 3 equal divisions and the interval on the $X$-axis into $3^2$ equal parts, as indicated in Fig. 31.

As $x$ increases through the first division, let $y$ increase from zero to $\frac{1}{3}$. As $x$ increases through the second division, let $y$ decrease from $\frac{1}{3}$ to zero. Then as $x$ increases through the third division, let $y$ again increase from zero to $\frac{1}{3}$. As $x$ goes through succeeding divisions, let $y$ increase or decrease as indicated in Fig. 31. If we indicate by $\Delta x$ the increments of $x$ by this division and by $\Delta y$ the corresponding increments of $y$, we have for the first division

$$\Delta x = \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2}, \frac{1}{3^2},$$

$$\Delta y = \frac{1}{3}, \frac{-1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{-1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{-1}{3}, \frac{1}{3}.$$

Divide each of the foregoing subintervals on the $Y$-axis into 3 equal parts and the subintervals on $X$-axis into $3^2$ equal parts. As $x$ increases through each of those divisions, let $y$ increase or decrease as for the former subdivision. Thus, for the $n^{th}$ division, we then have

$$\Delta x = \frac{1}{3^{2n}}, \frac{1}{3^{2n}}, \frac{1}{3^{2n}}, \cdots, \frac{1}{3^{2n}},$$

$$\Delta y = \frac{1}{3^n}, \frac{-1}{3^n}, \frac{1}{3^n}, \cdots, \frac{1}{3^n}. \tag{1}$$

---

[1] See, Steinitz, *Math. Ann.* Vol. 52, pp. 58–69; also, E. H. Moore, *Trans. Amer. Math. Soc.*, Vol. 1, pp. 81–90.

It is to be noted that the denominators of $\Delta x$ are powers of $3^2 = 9$, while those of $\Delta y$ are powers of 3. The numerators of $\Delta y$ are all numerically one, each positive one being preceded or followed by a negative one. Assume the foregoing method of determining $\Delta x$, $\Delta y$ to be carried on indefinitely. By this process, the function is defined at each point of the sequence

$$\frac{a_1}{9}, \frac{a_1}{9} + \frac{a_2}{9^2}, \frac{a_1}{9} + \frac{a_2}{9^2} + \frac{a_3}{9^3}, \cdot \cdot \cdot , \frac{a_1}{9} + \frac{a_2}{9^2} + \cdot \cdot \cdot + \frac{a_n}{9^n}, \cdot \cdot \cdot , \tag{2}$$

where $a_1, a_2, \cdot \cdot \cdot , a_n, \cdot \cdot \cdot$ may have any of the values

$$0, 1, 2, \cdot \cdot \cdot , 8.$$

For convenience, we shall denote this set of points by $\{x_k\}$. As the numerators $a_k$ take all possible values, the points of this set are everywhere dense in the interval (0, 1), and consequently every point of the interval can be represented by a sequence of such points, when we include its limiting points, since it is equivalent to saying that any real number can be represented in a number system having the base 9, which, as we have seen, is always possible. At those points which are limiting points of the set $\{x_k\}$ and not themselves points of the set, we shall define the function $f(x)$ as the limit of the functional values in the neighborhood of the points in question. With this understanding the function is fully defined for the entire interval (0,1).

The function $f(x)$ thus defined is continuous in the given interval (0, 1). For, as will be seen from (1), the maximum oscillation of the function in any interval by the $n^{th}$ division, that is the maximum value of $\Delta y$, is $\frac{1}{3^n}$. As $n$ increases without limit, the subinterval $\Delta x$ approaches zero and the corresponding oscillation of the function also approaches zero. Consequently, the function is continuous in the entire interval (0, 1) as stated.

We shall now undertake to show that $f(x)$ has no derivative at any point of the interval (0, 1). Let $x_0$ be any point in the given interval. In order that the derivative shall exist at $x_0$ we must have the same limiting value of the difference quotient

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

as $x_1$, $x_2$ approach simultaneously the limiting value $x_0$ through all possible sets of points dense at $x_0$, where $x_1$ and $x_2$ are on opposite sides of $x_0$. Consequently, if we can find two such sets of points which lead to different limiting values, it follows that the derivative does not exist at $x_0$. For one set of points, we shall make use of the points $x_k$ already discussed. As we have seen, these points are everywhere dense, and hence $x_0$ may be made the limiting point of such a set. For the values of $(x_2 - x_1)$ we shall take those values of $\Delta x$ given in (1). The corresponding values of $[f(x_2) - f(x_1)]$ are then the values of $\Delta y$ given in (1). For this method of selection, we have

$$\underset{\substack{x_1\to x_0\\ x_2\to x_0}}{L}\left|\frac{f(x_2)-f(x_1)}{x_2-x_1}\right| = \underset{\Delta x\to 0}{L}\left|\frac{\Delta y}{\Delta x}\right| = \underset{n\to\infty}{L}\frac{\dfrac{1}{3^n}}{\dfrac{1}{3^{2n}}} = \underset{n\to\infty}{L}\, 3^n = \infty. \tag{3}$$

From this conclusion, it follows that $f(x)$ can not have a finite derivative at $x_0$.

It can also be shown that $f(x)$ can not have an infinite derivative at $x_0$; for, in order that this may be true, the difference quotient must become infinite with a definite sign as the points $x_2$, $x_1$ approach $x_0$ in any manner whatever. To show that an infinite derivative does not exist, it is sufficient to show that, for a set of values for $x_1$, $x_2$ different from those already used, the limiting value of the difference quotient is finite. We shall undertake to show that, for at least one such set of values, this limit is zero. To make clear how this may be done, suppose we denote the sequence of values of $\Delta y$ in (1) by

$$\{\Delta_{nk}\},\ k = 1, 2, 3, \cdot\ \cdot\ \cdot, 3^{2n}.$$

Thus, for $n = 1$, we have

$$\Delta_{11} = \frac{1}{3},\ \Delta_{12} = \frac{-1}{3},\ \Delta_{13} = \frac{1}{3},\ \Delta_{14} = \frac{1}{3},$$

$$\Delta_{15} = \frac{-1}{3},\ \Delta_{16} = \frac{1}{3},\ \Delta_{17} = \frac{1}{3},\ \Delta_{18} = \frac{-1}{3},\ \Delta_{19} = \frac{1}{3}.$$

It will be observed that of any three consecutive values of $\Delta_{1k}$, one is always of opposite sign to the other two. Consequently, by combining each of these values either with the following or with the preceding, the sum can always be made zero. The same is true of the values of the sequence $\{\Delta_{nk}\}$ obtained by the $n^{th}$ division. We shall make use of such a combination to obtain a second set of values for $\Delta y$, which we may, for convenience, denote by $\{\Delta'_{nk}\}$. Thus for $n = 1$, we have

$$\begin{array}{lll} \Delta'_{11} = \Delta_{11} + \Delta_{12}, & \Delta'_{12} = \Delta_{12} + \Delta_{13}, & \Delta'_{13} = \Delta_{13} + \Delta_{12}, \\ \Delta'_{14} = \Delta_{14} + \Delta_{15}, & \Delta'_{15} = \Delta_{15} + \Delta_{16}, & \Delta'_{16} = \Delta_{16} + \Delta_{15}, \\ \Delta'_{17} = \Delta_{17} + \Delta_{18}, & \Delta'_{18} = \Delta_{18} + \Delta_{19}, & \Delta'_{19} = \Delta_{19} + \Delta_{18}. \end{array}$$

The corresponding values of

$$\Delta x = x_2 - x_1$$

are all numerically equal to $\frac{2}{3^2}$, and those of the $n^{th}$ division are $\frac{2}{3^{2n}}$. The end-points of these divisions of the given interval are also everywhere dense in (0, 1) and $x_0$ may be made therefore the limiting point of these new points. It will be seen also that each $\Delta y \equiv f(x_2) - f(x_1)$ in the method of division given by (1) is always a subinterval of the corresponding value of $\Delta' y$. Moreover, as $\Delta y$ approaches zero, $\Delta' y$ also approaches zero; in fact, it is so taken that it is always exactly zero, before passing to the limit with $\Delta x$. We have then for the new set of values for $\Delta x$, $\Delta y$

$$\underset{\substack{x_2 \to x_0 \\ x_1 \to x_0}}{L} \left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right| = \underset{\Delta x \to 0}{L} \left| \frac{\Delta y}{\Delta x} \right| = \underset{n \to \infty}{L} \frac{0}{\frac{2}{3^{2n}}} = 0. \qquad (4)$$

Since the limiting values given by (3) and (4) are different, it follows that the derivative at $x_0$ does not exist. As $x_0$ was taken to be any point of the interval (0, 1), it follows that $f(x)$ can not have a derivative either finite or infinite at any point in the given interval.

In Art. 19 it has been shown that a one-to-one correspondence can be established between the points of a square and the points on one side. This relation involves a parametric representation of the functional relation between $x$ and $y$, say

$$x = \phi_1(t),\ y = \phi_2(t),$$

where $t$ takes all values between zero and one. So long as a one-to-one correspondence is involved, it can be shown that the functions $\phi_1$, $\phi_2$ can not both be continuous.[1] If we omit this condition so that to a point $(x, y)$ more than one value of $t$ may correspond, then it may be shown that $\phi_1$, $\phi_2$ may be continuous and uniquely determined functions. Indeed, these functions can be so defined that the corresponding curve will be continuous and pass through every point of the square

$$0 \leqq x \leqq 1,\ 0 \leqq y \leqq 1$$

---

[1] See Netto, *Journal für die reine und angewandte Math.*, Vol. 86 (1878), pp. 263–268.

at least once, and have no definite direction at any point; that is, $y$ is then a continuous function of $x$ and yet the derivative $\frac{dy}{dx}$ does not exist at any point on this space filling curve.[1]

**31. Properties of derivatives and derivative numbers.** Let us suppose that the function $f(x)$ is single-valued and continuous in a given interval $(a, b)$. We shall now consider some of the special properties of its derivative and derivative numbers. Among these properties, that given in the following theorem, known as the **Law of the Mean,** is of fundamental importance.

THEOREM I. *Let $f(x)$ be continuous in the closed interval $(a, b)$, and suppose that $f'(x)$ exists for $a < x < b$. Then there exists a point $x_0$, for which*

$$\frac{f(b) - f(a)}{b - a} = f'(x_0),\ a < x_0 < b.$$

The given hypothesis assumes the existence of $f'(x)$ merely for the inner points of the interval $(a, b)$; that is, for points of the open interval $(a, b)$. It may be finite or infinite with a definite sign. The given function $f(x)$ is continuous at the end-points of the interval as well as at the interior points. We shall assume that $f(x)$ is not a constant; for, in that case the theorem follows at once, since $f'(x)$ is then zero for all values of $x$. We shall also assume $f(b)$ different from $f(a)$; for, if $f(b)$ is equal to $f(a)$ we can always replace the given interval by another where this is not the case. Consider the auxiliary function

$$F(x) = f(b) - f(x) - \frac{f(b) - f(a)}{b - a}(b - x).$$

We have then the relation

$$F(b) = F(a) = 0.$$

Moreover, $F(x)$ is continuous in the closed interval $(a, b)$ and has a derivative at each point of the open interval $(a, b)$. Then by Theorem IV, Art. 26, there exists at least one point $x_0$ interior to $(a, b)$ at which $F(x)$ takes its maximum value. We must then have for all values of $\Delta x \neq 0$,

$$F(x_0 - \Delta x) - F(x_0) \leqq 0,\ F(x_0 + \Delta x) - F(x_0) \leqq 0.$$

---

[1] For discussion of such curves, see Peano, *Math. Annalen*, Vol. 36, pp. 157–160 (1890); Hilbert, *Math. Annalen*, Vol. 38, pp. 459–460 (1891); E. H. Moore, *Trans. Amer. Math. Soc.*, Vol. 1, pp. 72–80 (1900).

The difference quotients

$$\frac{F(x_0 - \Delta x_0) - F(x_0)}{-\Delta x}, \ \frac{F(x_0 + \Delta x) - F(x_0)}{\Delta x}, \ \Delta x \neq 0,$$

must be either both zero or of opposite signs. It follows that as $\Delta x$ approaches 0, the limiting values of these difference quotients, namely the left-hand derivative and the right-hand derivative, respectively, must be of opposite signs if both are different from zero. As the derivative of $F(x)$ exists at $x_0$, the left-hand and right-hand derivatives must be equal, and hence each equal to zero. Consequently, we have

$$F'(x_0) = -f'(x_0) + \frac{f(b) - f(a)}{b - a} = 0,$$

or

$$\frac{f(b) - f(a)}{b - a} = f'(x_0), \ a < x_0 < b,$$

as the theorem requires.

Among other properties of derivative functions, we have those given by the following theorems.

THEOREM [1] II. *If $f(x)$ is continuous and has a derivative at every point in an interval $(a, b)$, then $f'(x)$ can not pass from a value $A$ to a value $B$ without taking all intermediate values.*

Let $x_1$, $x_2$ be the values of $x$ in the given interval at which the derivative is $A$ and $B$, respectively; that is, let

$$f'(x_1) = A, \ f'(x_2) = B.$$

We may, without loss of generality, assume $B > A$. Let us consider first the case where one of these derivatives, say $f'(x_1)$, is positive and the other is negative. For convenience put

$$F(x, \Delta x) \equiv \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Since $f(x)$ is continuous, $F(x, \Delta x)$ is also continuous in $x$ for $\Delta x \neq 0$. Then there exists between $x_1$ and $x_2$ some point $x_0$ for which $F(x, \Delta x)$ takes its maximum value. It follows that for all values of $\Delta x$ in the deleted neighborhood of zero, we have

$$F(x_0, -\Delta x) > 0, \ F(x_0, \Delta x) < 0.$$

---

[1] Due to Darboux, *Ann. Ecol. Normale* (2), Vol. 4 (1875), p. 109.

Since the derivative $f'(x)$ exists for $x = x_0$, these two functional values must approach the common limit zero as $\Delta x$ approaches zero. That is, we have

$$f'(x_0) = 0.$$

In other words, as the derivative function $f'(x)$ passes from $f'(x_1)$ to $f'(x_2)$, it must take at some point the value zero.

We shall now consider the case where $f'(x)$ does not change sign in passing from the value $A$ to the value $B$. Let $C$ be any value between $A$ and $B$. We shall show that there is some value $\bar{x}$ between $x_1$ and $x_2$ for which $f'(\bar{x}) = C$. As $C$ lies between $f'(x_1)$ and $f'(x_2)$, it follows that the derivative of the function

$$\psi(x) \equiv f(x) - Cx$$

will have opposite signs at $x_1$ and $x_2$; for, when $f'(x)$ takes the value $A$, we have

$$\psi'(x) = f'(x) - C = A - C,$$

which is negative, and when $f'(x)$ has the value $B$, we obtain

$$\psi'(x) = f'(x) - C = B - C,$$

which is positive. It follows from the first part of this discussion that there must exist some point $\bar{x}$ between $x_1$ and $x_2$ at which $\psi'(\bar{x})$ is zero; that is, where

$$\psi'(\bar{x}) \equiv f'(\bar{x}) - C = 0,$$

and hence

$$f'(\bar{x}) = C.$$

This conclusion establishes the theorem.

It is of value to note that the difference quotient

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}$$

likewise takes all values between its upper and lower limits in any closed interval $(a, b)$. Because, for $\Delta x \neq 0$ the difference quotient is fully defined and is continuous in the two variables $x$, $\Delta x$. As we have seen, every such function takes all values between any two of the values which it assumes.

Because $f'(x)$ takes all values between any two of its values it does not follow that it is continuous. Moreover, $f'(x)$ may exist at every point in a given interval and still be discontinuous as the following example illustrates.

Ex. 1. Given the function $f(x) = x^2 \sin \frac{1}{x}$, for $x \neq 0$, where $f(0) = 0$. Consider the existence and continuity of $f'(x)$.

For $x \neq 0$, we have

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}.$$

For $x = 0$, we obtain

$$f'(0) = \underset{\Delta x \to 0}{L} \frac{f(0 + \Delta x) - f(0)}{\Delta x} = \underset{\Delta x \to 0}{L} \Delta x \sin \frac{1}{\Delta x} = 0.$$

Hence, $f'(x)$ is fully defined in any interval inclosing the origin, but $f'(x)$ is not continuous at the origin, since the limit

$$\underset{x \to 0}{L} f'(x) \equiv \underset{x \to 0}{L} \left(2x \sin \frac{1}{x} - \cos \frac{1}{x}\right)$$

does not exist.

If a function $f(x)$ is continuous and has a derivative at every point of an interval, the question arises as to the number and distribution of the points of discontinuity of $f'(x)$. This question is settled by the following theorem.

THEOREM III. *If $f(x)$ is continuous and has a derivative at every point of an interval $(a, b)$, then $f'(x)$ is at most point-wise discontinuous in this interval.*[1]

To prove this theorem, let us define an auxiliary function $F(x, \Delta x)$ of the two independent variables $x$ and $\Delta x$ as follows:

$$F(x, \Delta x) \equiv \frac{f(x + \Delta x) - f(x)}{\Delta x}, \text{ for } \Delta x \neq 0.$$

For $\Delta x = 0$, let

$$F(x, 0) = \underset{\Delta x \to 0}{L} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x)$$

for values of $x$ in the interval $(a, b)$. The function $F(x, \Delta x)$ is continuous in $x$ for $\Delta x \neq 0$ and hence satisfies the conditions of Theorem X of Art. 28. Consequently, the limiting function $f'(x)$ can be at most point-wise discontinuous.

As a consequence of this theorem, we have the following corollaries.

COR. 1. *If $f(x)$ is continuous and has a derivative at every point of an interval $(a, b)$, then the points of continuity of $f'(x)$ form a set which is everywhere dense in $(a, b)$ and has the cardinal number of the continuum.*

---

[1] This theorem and the corollaries to follow hold also if one replaces the term *derivative* by *derivative number*. See Lebesgue, *Leçons sur l'intégration*, p. 121, *et seq.*

This result follows at once from the fact that under the conditions stated the derived function $f'(x)$ is point-wise discontinuous.

COR. 2. *If $f(x)$ is continuous and has a derivative at every point of an interval $(a, b)$, then $f'(x)$ is a function of at most Baire's class one.*

This conclusion follows from the fact that $f'(x)$ is at most a point-wise discontinuous function and is the limit as $\Delta x$ approaches zero of the difference quotient

$$F(x, \Delta x) \equiv \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

which is continuous in $x$ for $\Delta x$ different from zero.

From the foregoing theorem, it also follows that if $f'(x)$ exists at every point of a given interval, then it can not become infinite at a set of points everywhere dense. If $f(x)$ is not a constant in a given interval, then $f'(x)$ can not be zero at every point, and moreover it follows that it can not be exclusively zero or infinite at all the points of the interval; for, in that case it would not be point-wise discontinuous.

As a condition for the continuity of $f'(x)$, we have the following theorem.

THEOREM IV. *If $f(x)$ is continuous and has a derivative at every point in a closed interval $(a, b)$ a necessary and sufficient condition that $f'(x)$ is continuous in $(a, b)$ is that the difference quotient*

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}$$

*converges uniformly.*

That the condition is sufficient follows at once from Theorem IX of Art. 26. For, putting as before

$$F(x, \Delta x) \equiv \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

we have a function $F(x, \Delta x)$ of the independent variables $x$ and $\Delta x$ satisfying the conditions of that theorem. Hence the limiting function $f'(x)$ is continuous in the given interval.

The given condition is also necessary. For, from the Law of the Mean, we have

$$F(x, \Delta x) = f'(x + \theta\Delta x), \quad 0 < \theta < 1. \tag{1}$$

As the limiting function $f'(x)$ is continuous in the closed interval $(a, b)$, it is uniformly continuous in that interval. Hence, for a given $\epsilon$ there exists a $\delta$ independent of $x$ such that for every value $x_0$ in $(a, b)$ we have

$$|f'(x) - f'(x_0)| < \epsilon, \quad |x - x_0| < \delta.$$

Replacing $x$ by $x + \theta\Delta x$, this inequality becomes

$$|f'(x + \theta\Delta x) - f'(x_0)| < \epsilon,$$

which holds for every $x$ and for all values of $\Delta x$ such that

$$|x + \theta\Delta x - x_0| < \delta.$$

It is therefore valid for

$$|x - x_0| < \frac{\delta}{2}, \quad |\Delta x| < \frac{\delta}{2}.$$

In other words, we have

$$\underset{\substack{x \to x_0 \\ \Delta x \to 0}}{L} f'(x + \theta\Delta x) = f'(x_0).$$

From (1) it then follows that for every $x_0$ in the closed interval $(a, b)$ we have

$$\underset{\substack{x \to x_0 \\ \Delta x \to 0}}{L} F(x, \Delta x) = f'(x_0),$$

and hence from Theorem VII, Art. 26 the difference quotient $F(x, \Delta x)$ converges uniformly to its limiting function $f'(x)$.

THEOREM V. *Let $f(x)$ be continuous in the closed interval $(a, b)$ and have a derivative at each point in the deleted neighborhood of $x_0$, where $a < x_0 < b$. Then, if $\underset{x \to x_0}{L} f'(x)$ exists, the derivative $f'(x_0)$ exists and*

$$f'(x_0) = \underset{x \to x_0}{L} f'(x).$$

Let $\Delta x$ be so chosen that $x_0$ and $x_0 + \Delta x$ both fall within $(a, b)$. By the Law of the Mean, we have

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0 + \theta\Delta x), \quad 0 < \theta < 1. \tag{1}$$

But by hypothesis $f'(x)$ has a limit as $x$ approaches $x_0$, and hence

$$\underset{\Delta x \to 0}{L} f'(x_0 + \theta\Delta x) = A.$$

Consequently, the left-hand member of (1) must also approach the limiting value $A$ as $\Delta x$ approaches zero. Hence, $f(x)$ has a derivative at $x_0$ and we may write

$$f'(x_0) = \underset{\Delta x \to 0}{L} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = A,$$

and the theorem is established.

The discussions of this article would not be complete without mentioning some of the properties of derivative numbers. Among these properties we have those given in the following theorems.

THEOREM VI. *If, in the interval* $(a, b)$, $M$ *and* $m$ *denote respectively the least upper bound and the greatest lower bound of any one of the four derivative numbers of the continuous function* $f(x)$, *it follows that*

$$m \leqq \frac{f(b) - f(a)}{b - a} \leqq M.$$

We shall demonstrate the theorem for the case of the upper right-hand derivative number, $D^+f(x)$. A similar proof holds for the other derivative numbers. To establish the second inequality in the theorem, it is sufficient to show that there exist values of $D^+f(x)$ at least as large as $\frac{f(b) - f(a)}{b - a}$.

Let $P$ and $Q$ be the points on the curve $y = f(x)$ corresponding respectively to $x = a$, $x = b$. If there are points on the curve above the line $PQ$, let $P'Q'$ be a straight line parallel to $PQ$ and cutting the given curve at points whose ordinates are greater than those of $P$ and $Q$. In this case it is always possible to find some point $A$ of intersection at which the value of $D^+f(x)$ is at least equal to the tangent of the angle which $PQ$ makes with the $X$-axis.

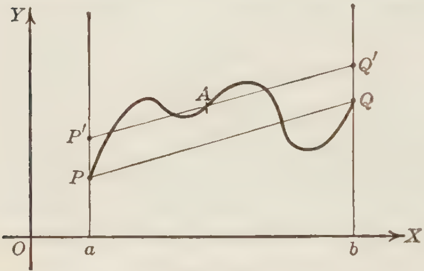


FIG. 32.

If the given curve has no points above $PQ$ but points lying below, it is always possible to find a line parallel to $PQ$ and lying below it so that the same condition holds at some point common to the given curve and the straight line.

A similar proof shows that $\frac{f(b) - f(a)}{b - a}$ must be equal to or be greater than $m$. If the given curve $y = f(x)$ coincides with $PQ$, the required property follows at once.

THEOREM VII. *For continuous functions, the four derivative numbers all have the same least upper bound and the same greatest lower bound in a given interval, and these upper and lower bounds are also the least upper and greatest lower bounds of the difference quotient*

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2},$$

*where* $x_1$, $x_2$ *take all possible values in the given interval.*

We shall establish the theorem for the derivative numbers $D_-f(x)$ and $D^+f(x)$. A similar proof holds for other possible cases and may be readily supplied.

Denote the least upper bound of $D^+f(x)$ in a given interval $(a, b)$ by $M$. It follows from the previous theorem that the difference quotient

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha},$$

can not exceed $M$, where $\alpha$ and $\beta$ are points in the given interval. We can, however, so select $\alpha$ and $\beta$ that this difference quotient shall be larger than $M - \epsilon$, where $\epsilon$ is arbitrarily small. The least upper bound of $D_-f(x)$ in such an interval $(\alpha, \beta)$, and hence in $(a, b)$, is also at least equal to $M - \epsilon$. As the values of $D_-f(x)$ can not exceed the least upper bound of the difference quotient, it follows that the two derivative numbers $D^+f(x)$, $D_-f(x)$ and the difference quotient have the same least upper bound $M$ in the interval $(a, b)$. A corresponding argument shows that $D^+f(x)$, $D_-f(x)$ have also the same greatest lower bounds.

The following corollaries follow at once from the foregoing theorem.

COR. 1. *If one of the derivative numbers of a continuous function has a positive lower bound, then the function is monotone and increasing.*

COR. 2. *If one of the derivative numbers of a continuous function has a negative upper bound, then the function is monotone and decreasing.*

COR. 3. *If $f(x)$ is continuous and has a derivative at each point, then the least upper and greatest lower bounds of the derivative $f'(x)$ in any interval are the same as those of the difference quotient.*

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

This corollary follows at once as a consequence of the foregoing theorem, since if $f'(x)$ exists at all points of a given interval, it must be identical in value with the derivative numbers of $f(x)$ for that interval.

If all four of the derivative numbers are bounded, then we have the following condition

$$\left|\frac{f(x_1) - f(x_2)}{x_1 - x_2}\right| < M,$$

where $M$ is finite and $x_1$, $x_2$ are any two points in the given interval. This condition is often spoken of as the **Lipschitz condition.**

THEOREM VIII. *If a continuous function $f(x)$ has a derivative number which is continuous at $x_0$, then all of the derivative numbers are continuous at $x_0$ and $f(x)$ has a derivative at that point.*

If one of the derivative numbers, say $D^+f(x)$, is continuous at $x_0$, then it must take at this point its maximum value, and this value must also be its minimum value at $x_0$. By Theorem VII all the derivative numbers have in any interval the same upper bound and the same lower bound. Suppose the interval in question to contain $x_0$ as an inner point. As the length of the interval approaches zero, this upper bound approaches the maximum value of the derivative numbers at $x_0$. In a similar manner the lower bound approaches the common minimum value of the derivative numbers at $x_0$. However, these values must be the same, since they are equal to those of $D^+f(x)$ and, as we have seen, the maximum and the minimum values of this derivative number are identical because of its continuity. As all of the derivative numbers take the same value at $x_0$, it follows that the derivative of $f(x)$ exists at that point.

It follows from this theorem that if one of the derivative numbers of a continuous function is zero for all values of $x$ in a given interval, then the function has a derivative everywhere equal to zero and, hence, the function is a constant.

The converse of this theorem is not true; that is, a function may have a derivative at every point of an interval and yet the derivative numbers may not be continuous. For example, let

$$\begin{aligned} f(x) &= x^2 \sin \frac{1}{x}, \quad x \neq 0, \\ &= 0 \qquad\qquad\;\, , \quad x = 0. \end{aligned}$$

The function has a derivative at every point in the interval (0, 1), namely

$$\begin{aligned} f'(x) &= 2x \sin \frac{1}{x} - \cos \frac{1}{x}, \quad x \neq 0, \\ &= 0 \qquad\qquad\qquad\qquad\;\, , \quad x = 0. \end{aligned}$$

However, none of the derivative numbers are continuous at $x = 0$, since they must all be equal to the derivative at any point and the derivative is discontinuous at the origin.

THEOREM IX. *A function is determined except as to an additive constant if we know a finite derivative number of the function for all values of the variable.*[1]

---

[1] See DuBois-Reymond, *Math. Annalen*, Vol. 16 (1880), p. 128.

This theorem is equivalent to saying that any two functions $f_1(x)$ and $f_2(x)$ which have the same finite derivative number for each value of the variable can differ at most by an additive constant. Let us assume, for example, that at each point of the given interval these two functions have the same upper right-hand derivative number; that is, that we have

$$D^+f_1(x) = D^+f_2(x). \tag{2}$$

From the geometric significance of derivative numbers, it follows that

$$D_+[-f_2(x)] = -D^+f_2(x). \tag{3}$$

From the definition of derivative numbers, we also have

$$D_+[f_1(x) - f_2(x)] \leqq D^+f_1(x) + D_+[-f_2(x)] \leqq D^+[f_1(x) - f_2(x)].$$

From (2) and (3) the middle term is zero and hence

$$D_+[f_1(x) - f_2(x)] \leqq 0 \leqq D^+[f_1(x) - f_2(x)].$$

It follows that the two right-hand derivative numbers of the function

$$F(x) \equiv [f_1(x) - f_2(x)],$$

if different from zero, can never have the same sign for any value of $x$. The function $F(x)$ can then never increase nor decrease and consequently must be a constant. If the two right-hand derivative numbers of $F(x)$ are zero, then again $F(x)$ is a constant. Hence, we have in either case

$$f_1(x) = f_2(x) + c.$$

As to the case where the given derivative number becomes infinite, the foregoing theorem gives no information. In this respect the statement corresponds to that which may be made concerning a derivative; that is, we can say that two functions having the same finite derivative can differ only by an additive constant, but if the condition of finiteness is omitted, no such statement can be made.

The foregoing theorem has been generalized by Scheeffer[1] in that he has shown that a continuous function is determined except as to an additive constant if one knows for each value of $x$, except at most for an enumerable set of points $E$, a finite derivative number of the given function. As a consequence it follows that a continuous function is determined as soon as one knows that its derivative, or a derivative number, at the irrational points exists and is finite. Moreover, it also follows that any two functions can differ at most by an additive constant if they have the same finite derivative number for all values of the variable with the exception of an enumerable set.

---

[1] See *Acta Math.*, Vol. 5 (1884–5), pp. 282–283.

**32. Definition of partial derivatives.** Suppose we have given a function $z = f(x, y)$ of the two independent variables $x$ and $y$. If we regard $x$ as a constant, then $f(x, y)$ may have a derivative with respect to $y$, which we call the partial derivative with respect to $y$ and denote it symbolically by

$$\frac{\partial f}{\partial y},\quad f_y'(x, y).$$

For the point $(x_0, y_0)$, such a partial derivative is defined by the limit

$$\underset{\Delta y\to 0}{L}\frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}. \tag{1}$$

Similarly, holding $y$ constant we have the partial derivative with respect to $x$, represented symbolically by

$$\frac{\partial f}{\partial x},\quad f_x'(x, y),$$

and defined at the point $(x_0, y_0)$ by the limit

$$\underset{\Delta x\to 0}{L}\frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}. \tag{2}$$

The derivative functions $f_x'(x, y)$, $f_y'(x, y)$ are in general functions of $x$ and of $y$ and hence may themselves have partial derivatives with respect to either $x$, or $y$. We are thus led to the higher partial derivatives

$$\frac{\partial^2 f}{\partial x^2},\ \frac{\partial^2 f}{\partial y^2},\ \frac{\partial^2 f}{\partial y\,\partial x},\ \frac{\partial^2 f}{\partial x\,\partial y},\ \text{etc.}$$

As in the case of the partial derivatives of the first order, the derivatives of the second order may be defined by means of limits as follows.

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \underset{\Delta x\to 0}{L}\frac{f_x'(x_0 + \Delta x, y_0) - f_x'(x_0, y_0)}{\Delta x}\\ &= \underset{\Delta_1 x\to 0}{L}\ \underset{\Delta_2 x\to 0}{L}\frac{f(x_0+\Delta_1 x+\Delta_2 x, y_0) - f(x_0+\Delta_1 x, y_0) - f(x_0+\Delta_2 x, y_0) + f(x_0, y_0)}{\Delta_1 x\cdot\Delta_2 x}\\ &= \underset{\Delta x\to 0}{L}\frac{f(x_0 + 2\Delta x, y_0) - 2f(x_0 + \Delta x, y_0) + f(x_0, y_0)}{(\Delta x)^2},\ \text{(If } f_x''(x_0, y_0) \text{ exists).}\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial y^2} &= \underset{\Delta y\to 0}{L}\frac{f_y'(x_0, y_0 + \Delta y) - f_y'(x_0, y_0)}{\Delta y}\\ &= \underset{\Delta_1 y\to 0}{L}\ \underset{\Delta_2 y\to 0}{L}\frac{f(x_0, y_0+\Delta_1 y+\Delta_2 y) - f(x_0, y_0+\Delta_1 y) - f(x_0, y_0+\Delta_2 y) + f(x_0 y_0)}{\Delta_1 y\ \Delta_2 y}\\ &= \underset{\Delta y\to 0}{L}\frac{f(x_0, y_0 + 2\Delta y) - 2f(x_0, y_0 + \Delta y) + f(x_0, y_0)}{(\Delta y)^2},\ \text{(If } f_y''(x_0, y_0) \text{ exists).}\end{aligned} \tag{4}$$

$$\frac{\partial^2 f}{\partial y \partial x} = \underset{\Delta y \to 0}{L} \frac{f_x'(x_0, y_0 + \Delta y) - f_x'(x_0, y_0)}{\Delta y}$$

$$= \underset{\Delta y \to 0}{L} \; \underset{\Delta x \to 0}{L} \frac{f(x_0+\Delta x, y_0+\Delta y) - f(x_0+\Delta x, y_0) - f(x_0, y_0+\Delta y) + f(x_0, y_0)}{\Delta y \; \Delta x}. \quad (5)$$

$$\frac{\partial^2 f}{\partial x \partial y} = \underset{\Delta x \to 0}{L} \frac{f_y'(x_0 + \Delta x, y_0) - f_y'(x_0, y_0)}{\Delta x}$$

$$= \underset{\Delta x \to 0}{L} \; \underset{\Delta y \to 0}{L} \frac{f(x_0+\Delta x, y_0+\Delta y) - f(x_0+\Delta x, y_0) - f(x_0, y_0+\Delta y) + f(x_0, y_0)}{\Delta y \; \Delta x}. \quad (6)$$

From these definitions, it will be seen that any discussion of the properties of these derivatives involves iterated limits as well as single limits. Corresponding definitions for other higher derivatives may be readily formulated.

In this connection it must be pointed out that the existence of the higher derivatives implies the existence of the corresponding derivatives of lower order. Thus in order that $f_{yx}''$ should exist at a point it is necessary that the partial derivative $f_x'$ should exist in the neighborhood of that point. It is possible, however, for the limit defining $f_{yx}''$ to exist without the partial derivative $f_x'$ existing as the following illustration shows. In such a case the higher derivative can not be said to exist.

**Ex. 1.** Given $f(x, y) = x \sin \frac{1}{x} + y \sin \frac{1}{y}$, $x \neq 0$, $y \neq 0$, where $f(0, y) = y \sin \frac{1}{y}$, $y \neq 0$; $f(x, 0) = x \sin \frac{1}{x}$, $x \neq 0$; $f(0, 0) = 0$. Examine the existence of $f_x'$ and $f_{yx}''$ at $x = 0$, $y = 0$.

We have $\quad f_x'(0, 0) = \underset{\Delta x \to 0}{L} \frac{f(\Delta x, 0) - f(0, 0)}{\Delta x} = \underset{\Delta x \to 0}{L} \sin \frac{1}{\Delta x}.$

Since this last limit does not exist, the derivative $f_x'(0, 0)$ can not be said to exist. However, the limit used in defining the second derivative $f_{yx}''$ does exist, for we have

$$\underset{\Delta y \to 0}{L} \; \underset{\Delta x \to 0}{L} \frac{f(\Delta x, \Delta y) - f(\Delta x, 0) - f(0, \Delta y) + f(0, 0)}{\Delta x \; \Delta y}$$

$$= \underset{\Delta y \to 0}{L} \; \underset{\Delta x \to 0}{L} \frac{\Delta x \sin \frac{1}{\Delta x} + \Delta y \sin \frac{1}{\Delta y} - \Delta x \sin \frac{1}{\Delta x} - \Delta y \sin \frac{1}{\Delta y}}{\Delta x \; \Delta y}$$

$$= \underset{\Delta y \to 0}{L} \; \underset{\Delta x \to 0}{L} \; 0 = 0.$$

In spite of the fact that this limit is zero, the derivative $f_{yx}''(0, 0)$ can not be said to exist, since $f_x'(0, 0)$ does not exist.

**33. Properties of partial derivatives.** If the two partial derivatives $f_x'$, $f_y'$ exist and are finite at every point in a given region $R$, it then follows that $f(x, y)$ is continuous in $x$ and in $y$ in this region.

Consequently, $f(x, y)$ must be continuous with respect to $x$ and $y$ together at a set of points everywhere dense in $R$. That the existence of these two derivatives is not a sufficient condition that $f(x, y)$ is everywhere continuous in $(x, y)$ is shown by the following illustration.

**Ex. 1.** Given the function $f(x, y) = \frac{x^2y}{x^4 + y^2}$, for $x \neq 0$, $y \neq 0$, and let $f(0, 0) = 0$. Show that the partial derivatives $f_x'$, $f_y'$ exist everywhere in the region $-1 \leqq x \leqq 1$, $-1 \leqq y \leqq 1$, although $f(x, y)$ is discontinuous in $(x, y)$ at the origin.

For $x \neq 0$, $y \neq 0$, we have

$$f_x' = 2xy\frac{y^2 - x^4}{(x^4 + y^2)^2}, \quad f_y' = x^2\frac{x^4 - y^2}{(x^4 + y^2)^2}.$$

For $x = 0$, $y = 0$, we obtain

$$f_x' = \underset{\Delta x \to 0}{L}\frac{(\Delta x)\cdot 0}{(\Delta x)^4 + 0} = 0, \quad f_y' = \underset{\Delta y \to 0}{L}\frac{1}{\Delta y}\frac{0\cdot\Delta y}{0 + (\Delta y)^2} = 0.$$

Similarly, from the definitions of the last article, we have

$$f_x'(0, y) = f_x'(x, 0) = 0, \quad f_y'(0, y) = 0, \quad f_y'(x, 0) = \frac{1}{x^2}.$$

Consequently, the partial derivatives $f_x'$, $f_y'$ exist at all points of the given region. The function $f(x, y)$ is not continuous in $(x, y)$ at the origin as one sees by comparing the limiting values along the line $y = 0$, and along the curve $y = x^2$; for, in the first case we obtain the limiting value 0 and in the other case $\frac{1}{2}$.

The following theorem gives a sufficient condition for the continuity of the given function.

THEOREM I. *If $f_x'$, $f_y'$ both exist and are finite at each point of a given region and if one of these derivatives is bounded as to $(x, y)$, then $f(x, y)$ is continuous in $(x, y)$.*

Since both $f_x'$ and $f_y'$ exist and are finite, it follows that $f(x, y)$ is continuous in $x$ and in $y$ at every point of the given region. By hypothesis one of the given derivatives, say $f_x'$, is bounded for all values of $x$ and $y$ in the given region $R$. Let $y_0$ be any value of $y$ in $R$. From the continuity of $f(x, y)$ with respect to $y$, we have for any value of $x$ in $R$

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0).$$

For all values of $y \neq y_0$ and for all values of $x$, we have

$$| f_x'(x, y) | < M,$$

where $M$ is some finite number. Hence, from Theorem VIII, Art. 26, it follows that $f(x, y)$ converges uniformly to $f(x, y_0)$. A necessary and sufficient condition for the uniform convergence of a function $f(x, y)$ which is continuous in $x$ is that

$$\underset{\substack{x\to x_0\\ y\to y_0}}{L} f(x, y) = f(x_0, y_0),$$

where $x_0$ is any point on the line $y = y_0$ and lying in $R$. But $y_0$ is any value of $y$ in $R$; hence, this simultaneous limit exists for any point $(x_0, y_0)$ in $R$, and $f(x, y)$ is therefore continuous in $(x, y)$ as the theorem states.

We may state the following corollary of the above theorem.

COR. *If the partial derivatives* $f_x'$, $f_y'$ *exist and are continuous in* $(x, y)$ *in the neighborhood of* $(x_0, y_0)$, *then* $f(x, y)$ *is continuous in* $(x, y)$ *at* $(x_0, y_0)$.

This proposition follows as a consequence of the foregoing theorem; for, since the partial derivatives are continuous in $(x, y)$ in the neighborhood of $(x_0, y_0)$ there must exist about this point a closed region in which these derivatives are bounded in the two variables together. Hence, in the neighborhood of $(x_0, y_0)$ the conditions of the theorem are satisfied.

For convenience in the discussion of the theorems to follow, let us set up the auxiliary function $F(x, y, \Delta x)$ defined as follows. On the line $y = y_0$ consider the closed interval $(x_0 - \delta, x_0 + \delta)$. In this interval take any two points $x$ and $x + \Delta x$ such that in case $x \neq x_0$ they lie on opposite sides of $x_0$. Let us assume that $f(x, y)$ is continuous in $x$ and in $y$ and that for every value of $x$ the partial derivative $f_x'$ exists. We then put

$$F(x, y_0, \Delta x) \equiv \frac{f(x + \Delta x, y_0) - f(x, y_0)}{\Delta x}, \ \Delta x \neq 0.$$

Let the value of $F(x_0, y_0, 0)$ be defined by the relation

$$F(x_0, y_0, 0) \equiv \underset{\Delta x\to 0}{L} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x} = f_x'(x_0, y_0).$$

We shall now consider the oscillation of $F(x, y, \Delta x)$ in the neighborhood of $(x_0, y_0)$. If $M$ and $m$ denote respectively the least upper bound and the greatest lower bound of $F(x, y, \Delta x)$ in this neighborhood, then the oscillation $\omega$ is given by the relation

$$\omega = M - m.$$

It will be noted that $\omega$ varies with $\delta$ and indeed as $\delta$ increases $\omega$ never decreases, while for decreasing $\delta$ we have

$$\underset{\delta\to 0}{L}\ \omega(\delta) = 0;$$

for, as $\delta$ approaches zero $\Delta x$ also approaches zero, and we have the same limiting values for $M$ and $m$, namely $f_x'(x_0, y)$.

For a given value of $\omega$, say $\omega = \eta$, there exists for the point $(x_0, y_0)$ a definite interval on the line $y = y_0$, namely

$$\{x_0 - \delta(x_0, y_0),\ x_0 + \delta(x_0, y_0)\},$$

where $\delta(x_0, y_0)$ depends upon the point $(x_0, y_0)$ and is to be considered as the least upper bound of those values of $\delta$ for which the value of $\omega$ is equal to or less than $\eta$. We may now prove the following theorem.[1]

THEOREM II. *For a given value of $\omega$ there exists a neighborhood $(x_0 - \sigma, x_0 + \sigma;\ y_0 - \sigma, y_0 + \sigma)$ of the point $(x_0, y_0)$ in which $\delta(x, y)$ is an upper semi-continuous function; that is,*

$$\delta(x, y) < \delta(x_0, y_0) + \epsilon,$$

*where $\epsilon$ is an arbitrarily small positive number.*

By hypothesis the value of the oscillation $\omega$ in the interval $\{x_0 - \delta(x_0, y_0), x_0 + \delta(x_0, y_0)\}$ taken on the line $y = y_0$ is a given number $\eta$. If the interval is increased by $\epsilon$ in length by replacing $\delta(x_0, y_0)$ by $\delta(x_0, y_0) + \frac{\epsilon}{2}$, we then have an interval

$$\left\{x_0 - \delta(x_0, y_0) - \frac{\epsilon}{2}, x_0 + \delta(x_0, y_0) + \frac{\epsilon}{2}\right\}$$

in which the value of $\omega$ is greater than $\eta$, say equal to $\eta + \alpha$. Hence, we can select two points $x_1$ and $x_2$, distinct from $x_0$, in the original interval and two values of $\Delta x$ such that $0 < |\Delta x| < \delta(x_0, y_0) + \frac{\epsilon}{2}$, say $\Delta_1 x$ and $\Delta_2 x$, for which we have

$$|\,F(x_1, y_0, \Delta_1 x) - F(x_2, y_0, \Delta_2 x)\,| > \eta + \alpha_1, \tag{1}$$

where $\alpha_1 < \alpha$. One of the points $x_1 + \Delta_1 x$, $x_2 + \Delta_2 x$, say $x_1 + \Delta_1 x$, must lie exterior to the interval $\{x_0 - \delta(x_0, y_0), x_0 + \delta(x_0, y_0)\}$, while the other may be interior to this interval. Let $(x_0 - \gamma, x_0 + \gamma)$ be an interval on $y = y_0$ such that all of the points $x_1$, $x_2$, $x_1 + \Delta_1 x$,

---

[1] Cf. Baire, *Annali di Mat.* series III$^a$, Vol. III (1899), p. 105.

$x_2 + \Delta_2 x$ lie exterior to it. As $F(x, y, \Delta x)$ is continuous in $y$, it follows that for any value of $y$ sufficiently close to $y_0$, say such that $| y - y_0 | < \beta$, we have from (1)

$$| F(x_1, y, \Delta_1 x) - F(x_2, y, \Delta_2 x) | > \eta. \tag{2}$$

Denote by $\sigma$ the smallest of the three numbers $\beta$, $\gamma$, $\frac{\epsilon}{2}$ and consider the points $(x, y)$ within the square defined by the inequalities

$$x_0 - \sigma < x < x_0 + \sigma,\ y_0 - \sigma < y < y_0 + \sigma. \tag{3}$$

Through any such point draw a line $L$ parallel to the $X$-axis, and upon this line take the interval $\left\{ x_0 - \delta(x_0, y_0) - \frac{\epsilon}{2}, x_0 + \delta(x_0, y_0) + \frac{\epsilon}{2} \right\}$. Whatever value of $y$ is chosen, this interval contains the four points

$$(x_1, y),\ (x_2, y),\ (x_1 + \Delta_1 x, y),\ (x_2 + \Delta_2 x, y), \tag{4}$$

and hence from (2) the value of $\omega$ in this interval must be greater than $\eta$. To each point $(x, y)$ there corresponds a definite $\delta$ such that on the line $L$ there exists an interval $(x - \delta, x + \delta)$ in which $\omega$ is less than $\eta$.

Denote by $\delta(x, y)$ the least upper bound of all those values of $\delta$ for which the inequalities (3) hold. We must then have

$$\delta(x, y) < \delta(x_0, y_0) + \epsilon;$$

for, otherwise the interval $(x - \delta, x + \delta)$ would contain the points (4) and hence $\omega$ could not be less than $\eta$. Hence, $\delta(x, y)$ is upper semi-continuous.

In case of functions of a single variable, it was shown that the derivative $f'(x)$ is at most point-wise discontinuous. The corresponding theorem with respect to the partial derivatives of $f(x, y)$ can be stated as follows.

THEOREM III. *Given a function $f(x, y)$ which is continuous in $x$ and in $y$. If the partial derivative $f_x'$ (or $f_y'$) exists at every point in a given closed region $R$, then it is at most point-wise discontinuous in $(x, y)$.*

The value of $\delta(x, y)$ in general depends upon the choice of $\eta$, where these symbols are used as defined in the foregoing discussion. By Theorem II, $\delta(x, y)$ is an upper semi-continuous function of $(x, y)$. It follows, as in the discussion of uniform continuity, that within any given closed region there exists a subregion in which the greatest lower bound of $\delta$ is positive, however small $\eta$ may be chosen. Hence, there exist in any given closed region points where $\delta$ has this greatest lower bound however small the oscillation of $F(x, y, \Delta x)$ and consequently of $f_x'(x, y)$ may be.

Let $R$ be any region in which the greatest lower bound of $\delta(x, y)$ is $\delta_1 > 0$. We shall now show that at any point $P_0 \equiv (x_0, y_0)$ of this region the oscillation of $f_x'(x, y)$ with respect to $x$ and $y$ together is never greater than $2\eta$. Denote by $R_1$ a rectangle lying within $R$ and so chosen that it contains the point $P_0$ and its sides are parallel to the axes of coördinates, the dimension parallel to the $X$-axis being less than $\delta_1$. On the line $y = y_0$ select two points $A_0 \equiv (\bar{x}_0, y_0)$, $B_0 \equiv (\bar{x}_0, + \Delta x, y_0)$ lying within $R$, but on opposite sides of $P_0$. From the existence of $f_x'(x, y)$ at $P_0$, we have

$$| F(\bar{x}_0, y_0, \ \Delta x) - f_x'(x_0, y_0) | \leqq \eta. \tag{5}$$

Through $A_0$, $B_0$ draw two lines $L_1, L_2$ parallel to the $Y$-axis. On each of these lines select an arbitrarily small interval inclosing the points $A_0$, $B_0$, respectively. Let $L$ be a line parallel to the $X$-axis and sufficiently close to $y = y_0$ so that it intersects $L_1$, $L_2$ within the above intervals. Let $A \equiv (\bar{x}_0, y)$, $B \equiv (\bar{x}_0 + \Delta x, y)$ be the points of intersection. Because of the continuity of $f(x, y)$ in $y$, we have

$$| F(\bar{x}_0, y, \ \Delta x) - F(\bar{x}_0, y_0, \ \Delta x) | < \epsilon, \tag{6}$$

where $\epsilon$ is an arbitrarily small, positive number. Let $P$ be any point on the line $L$ between $A$ and $B$. We have then an arbitrarily small area about $P_0$ but containing the point $P$, such that we have, since $f_x'(x, y)$ exists at $P$, for all values of $y$ lying in this area

$$| f_x'(x, y) - F(\bar{x}_0, y, \ \Delta x) | \leqq \eta. \tag{7}$$

Upon combining (5), (6) and (7), we have

$$| f_x'(x, y) - f_x'(x_0, y_0) | < 2\eta + \epsilon.$$

However, $(x_0, y_0)$ was any point of $R$. As we have seen, there exist points in every region $R$ for which $\delta(x, y)$ has its positive lower bound, however small $\eta$ may be chosen. At every such point the oscillation of $f_x'(x, y)$ must be less than the arbitrarily small choice of $2\eta + \epsilon$, and hence $f_x'(x, y)$ must be continuous. As such points occur in every region, the points of continuity of $f_x'(x, y)$ are everywhere dense, and consequently it can be at most point-wise discontinuous in $(x, y)$, as the theorem states.

In a similar manner, it may be shown that $f_y'(x, y)$ is at most point-wise discontinuous.

**34. Total differentiability.** Suppose we have given a single-valued function $z = f(x, y)$ defined for a region $R$ given by the inequalities

$$a < x < b, \ \ c < y < d.$$

We shall inquire into the conditions under which we may express the total differential of the function $f(x, y)$ by means of the relation

$$dz = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy, \tag{1}$$

or what is the same thing, what conditions if any must be placed upon the partial derivatives $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$ in order that this relation shall constitute a satisfactory definition of such a differential. In order that the definition be satisfactory, it must be such that when we assume a relation between the variables $x$ and $y$, say

$$y = \phi(x), \tag{2}$$

or if we express each variable in terms of a parameter, say

$$x = \psi_1(t), \ \ y = \psi_2(t), \tag{3}$$

we may find by aid of (1) the total derivatives of $z$ with respect to $x$ and with respect to $t$, namely:

$$\frac{dz}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\cdot\frac{dy}{dx}, \tag{4}$$

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial f}{\partial y}\cdot\frac{dy}{dt}. \tag{5}$$

Moreover, we should be able to get the same numerical value of $\dfrac{dz}{dx}$ by the use of (4) as by differentiating directly with respect to $x$ the transformed function

$$z = F(x) \equiv f\{x, \phi(x)\}.$$

That this is not always true if we assume in (1) merely the existence of the partial derivatives $f_x', f_y'$ is illustrated by the following example.

**Ex. 1.** Given $z = f(x, y) = \dfrac{x^2 y}{3x^2 + y^3}$ for $x \neq 0$, $y \neq 0$; and let $f(0, 0) = 0$.

The partial derivatives $f_x'(0, 0)$, $f_y'(0, 0)$ both exist, each being equal to zero. If we now assume

$$y = \varphi(x) = x$$

as the curve along which we propose to restrict the variation of $x$ and $y$, we have from (4), for $x = 0$, $y = 0$,

$$\frac{dz}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dx} = 0 + 0 \cdot 1 = 0.$$

On the other hand, putting $y = x$ in the given function, we have

$$z = \frac{x}{3 + x}, \quad \text{for } x \neq 0,$$
$$= 0, \text{ for } x = 0,$$

whence we have for $x = 0$

$$\frac{dz}{dx} = \underset{\Delta x \to 0}{L} \frac{f(\Delta x) - f(0)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{1}{\Delta x} \cdot \frac{\Delta x}{3 + \Delta x} = \frac{1}{3}.$$

Likewise, if we transform the given function by means of the relations given in (3), it does not necessarily follow that we get the same numerical value by differentiating the transformed function with respect to $t$ as is obtained from the given function by use of (5), if no restrictions as to continuity are placed upon $f_x'$, $f_y'$, as the following example shows.

**Ex. 2.** Given[1] $z = \sqrt{|xy|}$

The partial derivatives $f_x'(0, 0)$, $f_y'(0, 0)$ exist; for, we have

$$f_x'(0, 0) = \underset{\Delta x \to 0}{L} \frac{f(\Delta x, 0) - f(0, 0)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{0}{\Delta x} = 0,$$

$$f_y'(0, 0) = \underset{\Delta y \to 0}{L} \frac{f(0, \Delta y) - f(0, 0)}{\Delta y} = \underset{\Delta y \to 0}{L} \frac{0}{\Delta y} = 0.$$

Hence, (5) gives
$$\frac{dz}{dt} = 0 \cdot \frac{dx}{dt} + 0 \cdot \frac{dy}{dt} = 0.$$

On the other hand, if we transform the given functions by the relations

$$x = \psi_1(t) = t, \quad y = \psi_2(t) = t, \tag{6}$$

we have
$$z = \sqrt{|\psi_1(t) \cdot \psi_2(t)|} = \sqrt{t^2} = t,$$

and hence
$$\frac{dz}{dt} = 1.$$

Consequently, the relation (1) can not be said to give the total differential, since it does not give the proper value of the total derivative along the curve (6).

These illustrations show that any statement of the condition for total differentiability which involves merely the existence of the partial derivatives $f_x'$, $f_y'$ is not sufficient; for, under any such conditions the relations (1), (2), (3) are not necessarily consistent.

---

[1] See Stolz, *Differential = und Integralrechnung*, p. 133.

For the purposes of elementary analysis it is sufficient, however, and often convenient to assume that these partial derivatives are continuous with respect to the two variables $(x, y)$ together. This assumption enables us to show by aid of the Law of the Mean the existence of the total differential as given by (1) and consequently the total derivatives in a particular direction as given by (4) and (5). As we shall see later, such an assumption as to continuity is broader than it need be.

A suitable condition for total differentiability may be set up by following the analogous condition for differentiability of a function of a single variable. The function $f(x)$ is differentiable at $x_0$ if there exists a number $A$ such that we have

$$\underset{\Delta x \to 0}{L} \frac{f(x_0 + \Delta x) - f(x_0) - A \cdot \Delta x}{\Delta x} = 0.$$

Here, $|\Delta x|$ is the distance between the fixed point $x_0$ and the variable point $x_0 + \Delta x$. For the case of two variables the distance between the fixed point $(x_0, y_0)$ and the variable point $(x_0 + \Delta x,\ y_0 + \Delta y)$ is $\sqrt{(\Delta x)^2 + (\Delta y)^2}$, which may be regarded as the simultaneous increment of the two variables and denoted briefly by $\Delta(x, y)$. The corresponding increment of the given function is then $f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$. The function $f(x, y)$ may then be said to be totally differentiable if there exist two numbers $A$ and $B$ such that we have the simultaneous double limit

$$\underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) - A \cdot \Delta x - B\Delta y}{\Delta(x, y)} = 0.$$

This form of the condition for total differentiability is that introduced by several writers.[1] As Fréchet points out, one might make use of what he calls the **écart**, namely $|\Delta x| + |\Delta y|$, instead of the distance function $\Delta(x, y)$ employed in the foregoing definition.

If the given function is totally differentiable, it follows from this definition that the partial derivatives $f_x'$ and $f_y'$ exist and are finite; for, putting $\Delta x$ equal to zero and passing to the limit as $\Delta y$ approaches zero, we see that $B$ is nothing else than the partial derivative

---

[1] See Fréchet *Nouvelles Annales de Math.* (1912), Vol. 71, p. 389; also Kowalewski, *Komplexen Veränderlichen* (1911), p. 186.

A similar definition was also made by Thomae as early as 1875 in his *Theorie der Bestimmten Integrale,* p. 37.

$f_y'(x_0, y_0)$. Likewise, putting $\Delta y$ equal to zero and passing to the limit, we see that $A$ is the partial derivative $f_x'(x_0, y_0)$. Hence the above definition may be written

$$\underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \frac{f(x_0+\Delta x,\ y_0+\Delta y)-f(x_0,\ y_0)-\Delta x\, f_x'(x_0,\ y_0)-\Delta y\, f_y'(x_0,\ y_0)}{\Delta(x,\ y)} = 0. \quad (7)$$

The reason why the numerical result obtained from equations (4) and (5), when we assume merely the existence of the partial derivatives $f_x'$, $f_y'$, does not always agree with the derivative obtained from the transformed function is readily seen from geometric considerations. The partial derivatives $f_x'(x_0, y_0)$, $f_y'(x_0, y_0)$ determine the slope of two tangent lines, lying respectively in planes parallel to the $(X, Z)$-coördinate plane and the $(Y, Z)$-coördinate plane, with respect to the projection of these lines upon the $(X, Y)$-plane. These two tangent lines determine a plane passing through the point $(x_0, y_0, z_0)$. On the other hand, if we assume a relation between $x$ and $y$, say those given by (2) or (3), and take the derivative of the corresponding transformed function we get the limiting position of a tangent line to the surface $z = f(x, y)$ as the point of tangency approaches $(x_0, y_0, z_0)$ along the curve corresponding to the assumed relation between $x$ and $y$. This tangent line may or may not lie in the plane through $(x_0, y_0, z_0)$ determined by $f_x'(x_0, y_0)$, $f_y'(x_0, y_0)$. This interpretation of the two results suggests a condition for the existence of total differentiability which may be stated as follows.

THEOREM I. *A necessary and sufficient condition that the function* $f(x, y)$ *is totally differentiable at* $(x_0, y_0)$ *is that the surface* $z = f(x, y)$ *has a tangent plane at the point* $(x_0, y_0, z_0)$ *which is not parallel to the Z-axis.*[1]

To show that this condition is necessary, we assume that the given function is totally differentiable at $(x_0, y_0)$. Hence, the partial derivatives $f_x'(x_0, y_0)$, $f_y'(x_0, y_0)$ exist and are finite. If we take any curve on the surface such that it passes through the point $(x_0, y_0, z_0)$ and has at this point a definite direction, then this curve has at that point a definite tangent, which is also tangent to the surface. If we can now show that all such tangents at $(x_0, y_0, z_0)$ lie in the plane determined by the tangents parallel respectively to the $(X, Z)$-plane and the $(Y, Z)$-plane, then that plane is the tangent plane at $(x_0, y_0, z_0)$.

[1] Cf. Fréchet, *Nouvelles Annales de Math.* (1912), Vol. 71, p. 436.

The tangent to a curve cut from the given surface by the plane $y = y_0$ is a line parallel to the $(X, Z)$-plane and is given by the equations

$$y = y_0, \quad z - z_0 = (x - x_0)\, f_x'(x_0, y_0).$$

The direction cosines of this tangent are therefore proportional to

$$1,\ 0,\ f_x'(x_0, y_0).$$

Similarly, the tangent line parallel to the $(Y, Z)$-plane is given by the equations

$$x = x_0, \quad z - z_0 = (y - y_0)\, f_y'(x_0, y_0),$$

and its direction cosines are proportional to

$$0,\ 1,\ f_y'(x_0, y_0).$$

The equation of the plane determined by these tangents is, therefore,

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ 1 & 0 & f_x'(x_0, y_0) \\ 0 & 1 & f_y'(x_0, y_0) \end{vmatrix} = 0;$$

that is,
$$(z - z_0) - (y - y_0)\, f_y'(x_0, y_0) - (x - x_0)\, f_x'(x_0, y_0) = 0. \qquad (8)$$

To obtain the equation of a tangent line to the surface at $(x_0, y_0, z_0)$, let us assume a second point $(x_0 + \Delta x, y_0 + \Delta y, z_0 + \Delta z)$ on the surface. The chord connecting these two points becomes in the limit, as $\Delta x$, $\Delta y$, $\Delta z$ approach zero, a tangent line to the surface. The direction cosines, $\cos \lambda$, $\cos \mu$, $\cos \nu$, of the chord are respectively proportional to

$$\frac{\Delta x}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}, \quad \frac{\Delta y}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}, \quad \frac{\Delta z}{\sqrt{(\Delta x)^2 + (\Delta y)^2}};$$

in fact, these expressions multiplied by $\sin \nu$ are the direction cosines. Upon passing to the limit, we have three numbers which are proportional to the direction cosines of a tangent line $L$ to the surface at $(x_0, y_0, z_0)$. As we shall see, these limiting values may be written

$$\cos \alpha,\ \cos \beta,\ \cos \alpha\, f_x'(x_0, y_0) + \cos \beta\, f_y'(x_0, y_0),$$

where $\alpha$, $\beta$ are the direction angles of the projection of $L$ on the $(X, Y)$-plane. For, if the direction angles of the projection of $L$ are $\alpha$, $\beta$ respectively, then from (7) we have

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\frac{f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)}{\Delta(x,y)}$$
$$=\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\left\{\frac{\Delta x}{\Delta(x,y)}\cdot f_x'(x_0,y_0)+\frac{\Delta y}{\Delta(x,y)}\cdot f_y'(x_0,y_0)\right\}$$
$$=\cos\alpha\cdot f_x'(x_0,y_0)+\cos\beta\cdot f_y'(x_0,y_0). \quad (9)$$

The equations of the tangent may then be written in the form

$$\frac{x-x_0}{\cos\alpha}=\frac{y-y_0}{\cos\beta}=\frac{z-z_0}{\cos\alpha\cdot f_x'(x_0,y_0)+\cos\beta\cdot f_y'(x_0,y_0)}. \quad (10)$$

Replacing $x-x_0$, $y-y_0$, $z-z_0$ in (8) by the numbers which by (10) are proportional to them, we see that (8) is satisfied, no matter what may be the values of $\alpha$, $\beta$. Hence, the tangent at $(x_0, y_0, z_0)$ lies in the plane determined by the tangents to the given surface at this point which are parallel respectively to the $(X, Z)$ and $(Y, Z)$-planes. This plane is then a tangent plane to the given surface. Moreover, since $f_x'(x_0, y_0)$, $f_y'(x_0, y_0)$ are both finite, it follows that this plane can not be parallel to the $Z$-axis.

We can show as follows that the condition given in the theorem is also sufficient for total differentiability. Assuming that the surface $z = f(x, y)$ has a tangent plane at $(x_0, y_0, z_0)$, we have for its equation

$$z-z_0=(y-y_0)\,f_y'(x_0,y_0)+(x-x_0)\,f_x'(x_0,y_0), \quad (11)$$

where $(x, y, z)$ are the coördinates of a variable point $P'$ on that plane. Replacing $(x-x_0)$, $(y-y_0)$, $(z-z_0)$ by $\Delta x$, $\Delta y$, $\Delta z$, respectively, we may write the foregoing equation in the form

$$\Delta z=\Delta x\,f_x'(x_0,y_0)+\Delta y\,f_y'(x_0,y_0),$$

where $\Delta z$ denotes the length of the perpendicular let fall from $P'$ upon the plane $S$ which passes through the point $Q \equiv (x_0, y_0, z_0)$ and is parallel to the $(X, Y)$-plane. The foot of this perpendicular is $P \equiv (x_0 + \Delta x, y_0 + \Delta y, z_0)$. Denote by $\Delta f$ the distance on this same perpendicular, extended if need be, measured from the point $P''$ where it cuts the given surface to the plane $S$. We have then

$$\Delta f=f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0);$$

and hence

$$\frac{\Delta f-\Delta z}{\sqrt{(\Delta x)^2+(\Delta y)^2}}$$
$$=\frac{f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)-\Delta x\,f_x'(x_0,y_0)-\Delta y\,f_y'(x_0,y_0)}{\Delta(x,y)}. \quad (12)$$

However, we may put

$$\Delta f = \sqrt{(\Delta x)^2 + (\Delta y)^2}\tan\phi_1,\ \ \Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2}\tan\phi_2,$$

where $\phi_1$, $\phi_2$ are the angles made respectively by the lines $P''Q$, $P'Q$ with $PQ$. Since the plane given by (11) is tangent to the surface, it follows that as $P''$ approaches $Q$ in any manner whatsoever; that is, as $\Delta x$, $\Delta y$ approach zero simultaneously, $\phi_1$, $\phi_2$ and hence $\tan\phi_1$, $\tan\phi_2$ approach the same limiting value. Consequently, we have

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\frac{\Delta f - \Delta z}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} = \underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\{\tan\phi_1 - \tan\phi_2\} = 0.$$

We have, therefore, from (12)

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\left\{\frac{f(x_0+\Delta x,\ y_0+\Delta y) - f(x_0,\ y_0) - \Delta x f_x'(x_0,\ y_0) - \Delta y f_y'(x_0,\ y_0)}{\Delta(x,\ y)}\right\} = 0;$$

that is, the given function is totally differentiable at $(x_0, y_0)$, as the theorem requires.

By introducing polar coördinates one may express analytically a necessary and sufficient condition for the existence of a total differential. Put

$$\Delta x = \rho\cos\theta,\ \ \Delta y = \rho\sin\theta.$$

From (7) we then have

$$\frac{1}{\rho}\{f(x_0+\rho\cos\theta,\ y_0+\rho\sin\theta) - f(x_0,\ y_0)\} - \cos\theta\, f_x'(x_0,\ y_0) - \sin\theta\, f_y'(x_0,\ y_0).$$

A necessary and sufficient condition that (7) has the limit zero as $\Delta x$, $\Delta y$ approach zero simultaneously; that is, that $f(x, y)$ is totally differentiable at $(x_0, y_0)$, is that

$$\frac{1}{\rho}\{f(x_0 + \rho\cos\theta,\ y_0 + \rho\sin\theta) - f(x_0,\ y_0)\}$$

converges uniformly in the closed interval $(0, \pi)$ to the limit

$$\cos\theta\, f_x'(x_0,\ y_0) + \sin\theta\, f_y'(x_0,\ y_0)$$

as $\rho$ approaches zero.

The definition of total differentiability given by (7) can now be expressed in the following equivalent form:

$$f(x_0 + \Delta x,\ y_0 + \Delta y) = f(x_0,\ y_0) + \Delta x\, f_x'(x_0,\ y_0) + \Delta y\, f_y'(x_0,\ y_0) + \Delta(x,\ y)\cdot E(\Delta x,\ \Delta y), \quad (13)$$

where

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L} E(\Delta x,\ \Delta y) = 0.$$

A definition somewhat similar in form is frequently employed.[1] Any one of these equivalent definitions renders the relations given in (1), (4), and (5) consistent with one another and with the results obtained by first transforming the given function by means of an assumed relation between the variables and then computing the derivative from the resulting function.

There is a certain advantage in the differential notation used in (1) to express the results of total differentiation. The relations expressed by (4) and (5) imply a knowledge of a relation between the variables $x$ and $y$; for, in (4) we have involved the derivative $\frac{dy}{dx}$ and in (5) the derivatives $\frac{dx}{dt}, \frac{dy}{dt}$. In other words, these relations give the results of differentiating in a given direction. On the other hand, the differential notation employed in (1) does not involve any condition as to direction.

We shall now consider some of the consequences of total differentiability and also some of the conditions that must be placed upon the partial derivatives $f'_x$, $f'_y$ in order that the function $f(x, y)$ may be totally differentiable.

As one of the consequences of total differentiability, we have the following theorem.

THEOREM II. *If* $z = f(x, y)$ *is totally differentiable at the point* $(x_0, y_0)$, *then the given function is continuous in* $(x, y)$ *together at* $(x_0, y_0)$.

This theorem follows at once from (7). In order that this limit shall be zero, it is necessary that the numerator shall have the limit zero. Since the last two terms in the numerator vanish as $\Delta x$, $\Delta y$ approach zero, we have the condition that

$$\underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0);$$

but this is the condition that $f(x, y)$ is continuous in $(x, y)$ together at $(x_0, y_0)$. Hence, the theorem follows.

From the foregoing theorem it follows that continuity in the two variables is a necessary condition for total differentiability. It is not, however, a sufficient condition, as the following illustration shows.

---

[1] See Stolz, *Grundzüge der Differential = und Integralrechnung*, Vol. I, p. 132; Pierpont, *Functions of Real Variables*, Vol. I, p. 268; Young, *Fundamental Theorems of Differential Calculus*, p. 21; *Encyclop. des Sci. Math.*, II 3, p. 287; *Nouvelles Annales de Math.*, Vol. 71 (1912) p. 391.

**Ex. 3.** Given the function $f(x, y) = \dfrac{xy}{\sqrt{(x^2+y^2)}}$ for $x \neq 0$, $y \neq 0$, and let $f(0, 0) = 0$.

The given function is continuous in $(x, y)$ at the origin; for, upon putting $x = \rho \cos \theta$, $y = \rho \sin \theta$, we have

$$\frac{xy}{\sqrt{(x^2+y^2)}} = \rho \cos \theta \sin \theta = \frac{\rho}{2} \sin 2\theta,$$

and this last expression converges uniformly to the value zero as $\rho$ approaches the limiting value zero. But the given function is not totally differentiable at the origin; for, since $f_x'(0, 0) = 0$, $f_y'(0, 0) = 0$, we have

$$\underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \frac{f(\Delta x, \Delta y) - f(0, 0) - \Delta x\, f_x'(0, 0) - \Delta y\, f_y'(0, 0)}{\Delta(x, y)}$$

$$= \underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \frac{\Delta x \Delta y}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} \cdot \frac{1}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} = \underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \frac{\Delta x \Delta y}{(\Delta x)^2 + (\Delta y)^2}.$$

This last limit does not exist, as we can see by comparing it with Ex. 4, Art. 23.

As we have seen in Art. 33, the existence of the partial derivatives $f_x'(x_0, y_0)$, $f_y'(x_0, y_0)$ ensures the continuity of $f(x, y)$ with respect to $x$ and with respect to $y$, but not with respect to $x$ and $y$ together.

The following theorem is a consequence of the definition set up for total differentiability and gives a sufficient although not a necessary condition for total differentiability.

THEOREM III. *If the partial derivatives $f_x'(x, y)$, $f_y'(x, y)$ exist at the point $(x_0, y_0)$ and if one of these derivatives exists for all values of $x, y$ in the neighborhood of $(x_0, y_0)$ and is continuous at that point in the two variables together, then $f(x, y)$ is totally differentiable at $(x_0, y_0)$.*

The first two terms in the numerator of (7) can be written in the form

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) + f(x_0 + \Delta x, y_0) - f(x_0, y_0).$$

Since by hypothesis the derivative $f_x'(x, y)$ exists at $(x_0, y_0)$, we have

$$f(x_0 + \Delta x, y_0) - f(x_0, y_0) = \Delta x\, f_x'(x_0, y_0) + \Delta x\, \epsilon(\Delta x),$$

where
$$\underset{\Delta x \to 0}{L} \epsilon(\Delta x) = 0.$$

We shall assume in accordance with the conditions of the theorem that $f_y'(x, y)$ exists for all values of $x$ and $y$ in the neighborhood of $(x_0, y_0)$ and is continuous in $(x, y)$ together at this point. By the Law of the Mean we then have

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) = \Delta y\, f_y'(x_0 + \Delta x, y_0 + \theta \Delta y)$$
$$0 < \theta < 1.$$

From the left-hand member of (7) we now have by aid of the foregoing relations

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\left[\frac{\Delta y}{\Delta(x,\,y)}\left\{f_y'(x_0+\Delta x,\;y_0+\theta\Delta y)-f_y'(x_0,\;y_0)\right\}+\frac{\Delta x}{\Delta(x,\,y)}\cdot\epsilon(\Delta x)\right]. \qquad (14)$$

Since $f_y'(x, y)$ is continuous in $(x, y)$ at $(x_0, y_0)$, we have

$$\underset{\substack{\Delta x\to 0\\ \Delta y\to 0}}{L}\left[f_y'(x_0+\Delta x,\;y_0+\theta\Delta y)\right]=f_y'(x_0,\;y_0).$$

Moreover, for $\Delta x\neq 0$, $\Delta y\neq 0$, we have

$$\left|\frac{\Delta x}{\Delta(x,\,y)}\right|<1,\quad\left|\frac{\Delta y}{\Delta(x,\,y)}\right|<1.$$

Consequently, the limit given in (14) exists and is zero; that is, the condition of total differentiability is satisfied. Hence, the theorem.

From this theorem it follows, of course, that a given function is totally differentiable if the two derivatives $f_x'$, $f_y'$ are continuous in $(x, y)$. This condition furnishes an easy and satisfactory test for the purposes of elementary analysis.

As we have seen, a function of a single variable may be continuous throughout an interval without having a derivative at any point of the given interval. Similarly, for functions of two independent variables it is possible for a function to be continuous with respect to both variables in a given region without being totally differentiable at any point of the given region.

From Theorem III of Art. 33, it follows that if we assume merely that $f_x'$, $f_y'$ exist, then each is continuous in $(x, y)$ together at a set of points which is at least everywhere dense and has the cardinal number of the continuum. As a consequence, we have the following theorem.

THEOREM IV. *Given a function $f(x, y)$ having the partial derivatives $f_x'$, $f_y'$ for all values of $(x, y)$ in the given region $R$. The points at which $f(x, y)$ is totally differentiable form a set which is at least everywhere dense in $R$ and has the cardinal number of the continuum.*

By the method of condensation of singularities to be considered in Chapter VII, it is possible, however, to set up a function which is totally differentiable at a set of points everywhere dense in the given region and at the same time fail to have a total differential at a set of points also everywhere dense in the same region. Thus it could have a total differential at all of the points $(x, y)$ where one or both the variables are irrational and fail to have this property where both are rational. It would not be possible, however, for the function to be totally differentiable merely at the points where both variables are rational. By aid of Theorem I this statement is equivalent to saying that if $f_x'$, $f_y'$ exist at every point in $R$, then the surface

$$z = f(x, y)$$

must have a tangent plane not parallel to the $Z$-axis at a set of points everywhere dense in $R$, but there may also exist a set of points everywhere dense at which the surface has no such tangent plane.

We have also the following theorem as a condition for total differentiability.

THEOREM V. *Given the function* $f(x, y)$ *whose partial derivatives* $f_x', f_y'$ *exist at* $(x_0, y_0)$; *then* $f(x, y)$ *is totally differentiable at* $(x_0, y_0)$ *if in the neighborhood of* $(x_0, y_0)$, $f(x, y)$ *is continuous in* $y$ *for* $x \neq x_0$ *and the difference quotient*

$$\frac{f(x_0 + \Delta x, y) - f(x_0, y)}{\Delta x} \tag{15}$$

*converges uniformly as to* $\Delta x$ *for all values of* $y$ *in that neighborhood.*

We set up the identity

$$\frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)}{\Delta(x, y)} - \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0 + \Delta y)}{\Delta(x, y)} - \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta(x, y)} = 0. \tag{16}$$

Since the partial derivative $f_y'(x_0, y_0)$ exists, we may write

$$\left| \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y} - f_y'(x_0, y_0) \right| < \epsilon, | \Delta y | < \delta_1$$

or
$$| f(x_0, y_0 + \Delta y) - f(x_0, y_0) - \Delta y \, f_y'(x_0, y_0) | < \epsilon \cdot | \Delta y | . \tag{17}$$

As the convergence of the difference quotient (15) is uniform for all values of $y$ in the neighborhood of $(x_0, y_0)$, it follows that for a given value of $\epsilon$ there exists a $\delta_0$ independent of $y$ such that

$$\left| \frac{f(x_0 + \Delta x, y) - f(x_0, y)}{\Delta x} - f_x'(x_0, y) \right| < \epsilon, | \Delta x | < \delta_0. \tag{18}$$

As this relation holds uniformly for all values of $y$ in the neighborhood of $y = y_0$, we may then write

$$\left|\frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0 + \Delta y)}{\Delta x} - f'_x(x_0, y_0 + \Delta y)\right| < \epsilon,$$
$$|\Delta x| < \delta_0, \; |\Delta y| < \delta_2. \qquad (19)$$

From the continuity of $f(x, y)$ as to $y$ and the uniform convergence of the difference quotient (15) it follows that $f'_x(x_0, y)$ is continuous in $y$ and hence

$$|f'_x(x_0, y_0 + \Delta y) - f'_x(x_0, y_0)| < \epsilon, \; |\Delta y| < \delta_3. \qquad (20)$$

Combining (19) and (20), we have

$$\left|\frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0 + \Delta y)}{\Delta x} - f'_x(x_0, y_0)\right| < 2\epsilon,$$
$$|\Delta x| < \delta_0, \; |\Delta y| < \delta,$$

where $\delta$ is the smaller of the two numbers $\delta_2$, $\delta_3$. From this inequality we have, after multiplying by $|\Delta x|$,

$$|f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0 + \Delta y) - \Delta x\, f'_x(x_0, y_0)| < 2\epsilon\, |\Delta x|, \qquad (21)$$

where $|\Delta x| < \delta_0$, $|\Delta y| < \delta$. Making use of the relation given by (17) and (21), we have from (16)

$$\left|\frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) - \Delta x\, f'_x(x_0, y_0) - \Delta y\, f'_y(x_0, y_0)}{\Delta(x, y)}\right|$$
$$< 2\epsilon \frac{|\Delta x| + |\Delta y|}{\Delta(x, y)} \leqq 2\sqrt{2}\cdot\epsilon.$$

Since $\epsilon$ is arbitrarily small, it follows that the left-hand member of this inequality approaches the limit zero as $\Delta x$, $\Delta y$ approach zero simultaneously. Hence, it follows that $f(x, y)$ is totally differentiable at $(x_0, y_0)$, as the theorem states.

A corresponding theorem may be stated by assuming that the difference quotient

$$\frac{f(x, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

converges uniformly. From the foregoing theorem, it also follows that $f(x, y)$ is totally differentiable in a given region $R$ if in this region $f'_x$, $f'_y$ exist and one of them converges uniformly. It is also sufficient if both of the derivatives $f'_x$, $f'_y$ exist and one of the second partial derivatives $f''_x$, $f''_y$ is bounded as to $(x, y)$ together; for, in this case, one of the first partial derivatives must converge uniformly.

**35. Interchange of the order of differentiation.** We shall inquire into the conditions under which

$$\frac{\partial^2 z}{\partial y\,\partial x} = \frac{\partial^2 z}{\partial x\,\partial y},$$

where $z = f(x, y)$ is a function of the two independent variables $x$ and $y$. That this is not always possible is illustrated by the following example.

**Ex. 1.** Given $f(x, y) = xy\left(\dfrac{x^2 - y^2}{x^2 + y^2}\right)$, for $x \neq 0$, $y \neq 0$, and let $f(0, 0) = 0$. Test the interchange of order of differentiation at the origin.[1]

For $x \neq 0$, $y \neq 0$, we have

$$\frac{\partial f}{\partial x} = y\left(\frac{x^2 - y^2}{x^2 + y^2} + \frac{4x^2y^2}{(x^2 + y^2)^2}\right), \quad \frac{\partial f}{\partial y} = x\left(\frac{x^2 - y^2}{x^2 + y^2} - \frac{4x^2y^2}{(x^2 + y^2)^2}\right).$$

To get the value of these derivatives for $x = 0$, $y = 0$, we must make use of the definition of a partial derivative rather than substituting the values of $x$ and $y$ in the above results; for, they were obtained on the condition that $x \neq 0$, $y \neq 0$, and are indeterminate for $x = 0$, $y = 0$. We have

$$f_x'(0, 0) = \underset{\Delta x \to 0}{L} \frac{\Delta x \cdot 0}{\Delta x}\left\{\frac{(\Delta x)^2 - 0}{(\Delta x)^2 + 0}\right\} = 0,$$

$$f_y'(0, 0) = \underset{\Delta y \to 0}{L} \frac{0 \cdot \Delta y}{\Delta y}\left\{\frac{0 - (\Delta y)^2}{0 + (\Delta y)^2}\right\} = 0.$$

Both of the first partial derivatives $f_x'$, $f_y'$ exist everywhere and are continuous at the origin. Hence, their derivatives are given by the following limits.[2]

$$f_{xy}''(0, 0) = \underset{\Delta x \to 0}{L}\ \underset{\Delta y \to 0}{L} \left\{\frac{f(0 + \Delta x, 0 + \Delta y) - f(0 + \Delta x, 0) - f(0, 0 + \Delta y) + f(0, 0)}{\Delta x\,\Delta y}\right\}$$

$$= \underset{\Delta x \to 0}{L}\ \underset{\Delta y \to 0}{L} \left\{\frac{(\Delta x)^2 - (\Delta y)^2}{(\Delta x)^2 + (\Delta y)^2}\right\} = 1,$$

$$f_{yx}''(0, 0) = \underset{\Delta y \to 0}{L}\ \underset{\Delta x \to 0}{L} \left\{\frac{(\Delta x)^2 - (\Delta y)^2}{(\Delta x)^2 + (\Delta y)^2}\right\} = -1.$$

---

[1] See Stolz, *Differential-und Integralrechnung*, Vol. 1, p. 150.

[2] The earlier writers assumed the interchange of the order of differentiation as evident. Nicholas Bernouilli, Euler, and Cauchy took this point of view. Doubtless P. H. Blanchet was the first [*Jour. Math.* (1) Vol. 6 (1841)] to examine the conditions under which the order of limits can be reversed, although his work was not rigorous, as may also be said of the investigations of the subject by Linderhöf [*Acta. Soc. Sci. Fennicae*, Vol. 8 (1867), No. 7] and Genocchi [*Atti. Acad. Torino*, Vol. 4 (1868–9)]. Hermite in his *Cours d'analyse* (1873) and Laurent in his *Traité d'analyse* (1885), following the method of Lagrange, make use of Taylor's series, thus assuming the existence of all of the partial derivatives. Later contributions in restricting the conditions imposed upon $f(x, y)$ and its derivatives have been made by Peano [*Mathesis* (1) Vol. 10, (1890)], Schwarz [*Math. Abh.*, Vol. 2, (1890)] and others.

Consequently, we have the condition that

$$\frac{\partial^2 f(0,0)}{\partial x\,\partial y} \neq \frac{\partial^2 f(0,0)}{\partial y\,\partial x},$$

although it may be shown that the order of differentiation can be interchanged at every other point of the finite region.

The following theorem gives a convenient test.

THEOREM I. *If the partial derivatives* $f_y'$ *and* $f_{yx}''$ *exist at all points in the neighborhood of* $(x_0, y_0)$ *and* $f_{yx}''$ *is continuous in* $(x, y)$ *at* $(x_0, y_0)$*, then* $f_{xy}''(x_0, y_0)$ *exists and is equal to* $f_{yx}''(x_0, y_0)$.

If we regard $y_0$ and $\Delta y$ as constant, we may put

$$F(x) = f(x, y_0 + \Delta y) - f(x, y_0).$$

From the existence of $f_{yx}''$, it follows that $f_x'$ must exist at all points in the neighborhood of $(x_0, y_0)$ and hence that $F(x)$ is continuous and $F'(x)$ must exist. By applying the Law of the Mean to $F(x)$ we have

$$F(x_0 + \Delta x) - F(x_0) = \Delta x\, F'(x_0 + \theta_1 \Delta x),\ 0 < \theta_1 < 1.$$

If written in terms of $f(x)$, this relation becomes

$$\{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0)\} - \{f(x_0, y_0 + \Delta y) - f(x_0, y_0)\}$$
$$= \Delta x\{f_x'(x_0 + \theta_1\Delta x, y_0 + \Delta y) - f_x'(x_0 + \theta_1 \Delta x, y_0)\}. \quad (1)$$

Regarding $x_0 + \theta_1\Delta x$ as a constant the right-hand member of this equation becomes by a second application of the Law of the Mean

$$\Delta x \Delta y\{f_{yx}''(x_0 + \theta_1\Delta x, y_0 + \theta_2\Delta y)\},\ 0 < \theta_2 < 1.$$

Hence, dividing by $\Delta x \Delta y$ we have

$$\frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) - f(x_0, y_0 + \Delta y) + f(x_0, y_0)}{\Delta x\,\Delta y}$$
$$= f_{yx}''(x_0 + \theta_1\Delta x, y_0 + \theta_2\Delta y).$$

As $f_{yx}''$ is continuous with respect to $x$ and $y$ together at $(x_0, y_0)$, we write

$$\underset{\substack{\Delta x \to 0\\ \Delta y \to 0}}{L} \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) - f(x_0, y_0 + \Delta y) + f(x_0, y_0)}{\Delta x\,\Delta y}$$
$$= \underset{\substack{\Delta x \to 0\\ \Delta y \to 0}}{L} f_{yx}''(x_0 + \theta_1\Delta x, y_0 + \theta_2\Delta y) = f_{yx}''(x_0, y_0).$$

Since, $f_y'$ exists at all points in the neighborhood of $(x_0, y_0)$ the limit of the first member must exist as $\Delta y$ approaches zero for each value of $\Delta x$ in the neighborhood of zero. Likewise, from the existence of $f'_x$, the limit of this expression must exist as $\Delta x$ approaches zero for each value of $\Delta y$ in the neighborhood of zero. Hence, by Theorem V, Art. 24, we may replace the double limit in the first member by the sequential limit taken in either order and hence write

$$\underset{\Delta x \to 0}{L} \; \underset{\Delta y \to 0}{L} \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) - f(x_0, y_0 + \Delta y) + f(x_0, y_0)}{\Delta x \, \Delta y} = f''_{yx}(x_0, y_0),$$

or

$$f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0),$$

which was to be established.

The conditions given in the theorem are sufficient but not necessary. For, as the following example shows, it is possible to interchange the order of differentiation even though neither $f''_{xy}$ nor $f''_{yx}$ is continuous in $(x, y)$ at the point.

**Ex. 2.** Given $f(x, y) = \frac{x^2 + y^2}{4} \log (x^2 + y^2)$, where $f(0, 0) = 0$. Discuss the derivatives $f''_{xy}(0, 0)$, $f''_{yx}(0, 0)$.

For $x \neq 0$, $y \neq 0$, we have

$$f'_x = \frac{x}{2}\left\{ 1 + \log(x^2 + y^2) \right\}, \quad f'_y = \frac{y}{2}\left\{ 1 + \log(x^2 + y^2) \right\},$$

and hence

$$f''_{xy} = f''_{yx} = \frac{xy}{x^2 + y^2}.$$

For $x = 0$, $y = 0$, we have from the definition of partial derivatives

$$f'_x = f'_y = f''_{xy} = f''_{yx} = 0.$$

While the order of differentiation can be interchanged at every point, nevertheless neither of the derivatives $f''_{xy}$, $f''_{yx}$ is continuous in $(x, y)$ at the origin, since the limit of these derivatives as $x$ and $y$ approach zero simultaneously is not zero.

If $f'_x$ is continuous in $x$ and in $y$ and the partial derivative $f''_{yx}$ exists and is finite at all points in a given closed region $R$, then it follows that $f''_{yx}$ is at most point-wise discontinuous in $(x, y)$. If $f'_y$ also exists everywhere in $R$, then by the foregoing theorem, the points at which $f''_{xy}$ must exist and be equal to $f''_{yx}$ are everywhere dense in $R$. Moreover, these points form a set having the cardinal number of the continuum. A corresponding statement can be made concerning the partial derivatives $f'_y$, $f'_x$, $f''_{xy}$. This result also shows that it is impossible for the two partial derivatives $f''_{xy}$, $f''_{yx}$ to both exist at every point in $R$ and be distinct at all points. Moreover, it would be impossible to have both exist at all points and be equal merely at those points where both coördinates are rational.

Consider the three ratios

$$F(x, y, \Delta x, \Delta y) \equiv \frac{f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y) - f(x, y + \Delta y) + f(x, y)}{\Delta x \, \Delta y},$$

$$\frac{f_x'(x, y + \Delta y) - f_x'(x, y)}{\Delta y}, \quad \frac{f_y'(x + \Delta x, y) - f_y'(x, y)}{\Delta x}.$$

By an argument similar to that employed in the proof of Theorem VII of Art. 31, it may be shown that these three ratios must have the same least upper bounds and the same greatest lower bounds in a given region $R$, provided, of course, that $x$, $y$, $\Delta x$, $\Delta y$ are so taken that the points

$$(x, y), (x + \Delta x, y), (x, y + \Delta y), (x + \Delta x, y + \Delta y)$$

all lie in $R$. Hence, the derivative numbers of $f_x'$ with respect to $y$ and of $f_y'$ with respect to $x$ have the same least upper and the same greatest lower bounds. These derivative numbers are functions of $x$ and $y$. They have the same $(x, y)$ saltus, at each point. At any point where this saltus is zero; that is, where they are continuous functions of $(x, y)$, the partial derivatives $f_{xy}''$, $f_{yx}''$ exist and are equal. Consequently, if we assume that $f_{xy}''$, $f_{yx}''$ both exist at every point of $R$, then it follows that at any point where one of them is continuous in $(x, y)$ the other is also, and hence from Theorem I they are equal.[1] If one is discontinuous in $(x, y)$ the other is also, and they have the same $(x, y)$-saltus at the point. Thus, for the function considered in Ex. 1 the partial derivatives $f_{xy}''$, $f_{yx}''$ are continuous except at the origin.

For functions of a single variable, we saw that if $f'(x)$ exists for $x \neq x_0$ and has a limit as $x$ approaches $x_0$, then that limiting value is the value of $f'(x)$ at $x_0$. The corresponding theorem for functions of two variables may be stated as follows.

THEOREM II. *Let $f(x, y)$ be continuous and have the partial derivative $f_{yx}''$ at all points other than $(x_0, y_0)$. Also, let $f_x'(x_0, y_0)$ exist and*

$$\underset{\substack{x \to x_0 \\ y \to y_0}}{L} f_{yx}''(x, y) = A.$$

*Then $f_{yx}''$ exists at $(x_0, y_0)$ and is equal to $A$. Moreover, if $f_y'$ also exists in the neighborhood of $(x_0, y_0)$, then $f_{xy}''$ exists at $(x_0, y_0)$ and is equal to $f_{yx}''$.*

---

[1] See Schwarz, *Gesammelte Math. Abhandlungen*, Vol. 2, pp. 275–284; also, W. H. Young, *Fundamental Theorems of Differential Calculus*, p. 67.

For convenience, put

$$F(\Delta x, \Delta y) \equiv f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) \\ - f(x_0, y_0 + \Delta y) + f(x_0, y_0), \quad (1)$$

and
$$\phi(x, \Delta y) \equiv f(x, y_0 + \Delta y) - f(x, y_0). \quad (2)$$

Since $f''_{yx}$ exists for $x \neq x_0$, $y \neq y_0$, it follows that $f'_x$ must also exist in the open interval $(x_0, x_0 + \Delta x)$, and hence from the Law of the Mean, we have, regarding $\Delta y$ as a constant,

$$F(\Delta x, \Delta y) = \phi(x_0 + \Delta x, \Delta y) - \phi(x_0, \Delta y) = \Delta x\, \phi'_x(x_0 + \theta_1 \Delta x, \Delta y), \\ 0 < \theta_1 < 1. \quad (3)$$

By differentiating (2) we obtain

$$\phi'_x(x, \Delta y) = f'_x(x, y_0 + \Delta y) - f'_x(x, y_0).$$

If we apply the Law of the Mean to the second member of this equation, we may write

$$\phi'_x(x, \Delta y) = \Delta y\, f''_{yx}(x, y_0 + \theta_2 \Delta y), 0 < \theta_2 < 1,$$

whence from (3)

$$F(\Delta x, \Delta y) = \Delta x\, \Delta y\, f''_{yx}(x_0 + \theta_1 \Delta x, y_0 + \theta_2 \Delta y). \quad (4)$$

Introducing the auxiliary function

$$\psi(\Delta x, y) \equiv f(x_0 + \Delta x, y) - f(x_0, y),$$

we may write $\psi(\Delta x, y) = \Delta x\{f'_x(x_0, y) + \rho(\Delta x, y)\}$,

where $\underset{\Delta x \to 0}{L} \rho(\Delta x, y) = 0$

for any value of $y$ in the given region $R$. Whence, we have from (1)

$$F(\Delta x, \Delta y) = \psi(\Delta x, y_0 + \Delta y) - \psi(\Delta x, y_0) \\ = \Delta x\{f'_x(x_0, y_0 + \Delta y) - f'_x(x_0, y_0) + \rho(\Delta x, y_0 + \Delta y) \\ - \rho(\Delta x, y_0)\}. \quad (5)$$

Equating the two values of $F(\Delta x, \Delta y)$ given in (4) and (5), we have after dividing by $\Delta x\, \Delta y$

$$\frac{f'_x(x_0, y_0 + \Delta y) - f'_x(x_0, y_0)}{\Delta y} + \frac{\rho(\Delta x, y_0 + \Delta y)}{\Delta y} - \frac{\rho(\Delta x, y_0)}{\Delta y} \\ = f''_{yx}(x_0 + \theta_1 \Delta x, y_0 + \theta_2 \Delta y).$$

Since the simultaneous limit of the second member exists as $\Delta x$, $\Delta y$ approach zero and is equal to $A$, the same limit must exist for the first member. However, since the limit of the first member as $\Delta x$ approaches zero also exists for all values of $\Delta y \neq 0$, we may evaluate the simultaneous limit by first allowing $\Delta x$ to approach zero and then allowing $\Delta y$ to approach zero. Passing to the limit in this manner, we have

$$f''_{yx}(x_0, y_0) = \underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} f''_{yx}(x_0 + \theta_1 \Delta x, y_0 + \theta_2 \Delta x) \equiv \underset{\substack{x \to x_0 \\ y \to y_0}}{L} f''_{yx}(x, y) = A,$$

which establishes the first part of the theorem, and hence $f''_{yx}$ is continuous in both variables together at $(x_0, y_0)$.

If in addition to the existence and continuity of $f''_{yx}$ we have the condition that $f'_y$ exists in the neighborhood of $(x_0, y_0)$, then it follows from Theorem I that $f''_{xy}$ also exists at $(x_0, y_0)$ and is equal to $f''_{yx}(x_0, y_0)$.

Either of the derivatives $f''_{xy}$, $f''_{yx}$ can exist at a point $(x_0, y_0)$ even though the simultaneous limit of the derivative does not exist. The function considered in Ex. 1 furnishes a simple illustration. As we have seen, both of the partial derivatives $f'_x$, $f'_y$ exist at the origin. For $x \neq 0$, $y \neq 0$, we have

$$f''_{xy} = f''_{yx} = \frac{x^2 - y^2}{x^2 + y^2} \left\{ 1 + \frac{8x^2y^2}{(x^2 + y^2)^2} \right\}.$$

However, the limit

$$\underset{\substack{x \to 0 \\ y \to 0}}{L} \frac{x^2 - y^2}{x^2 + y^2} \left\{ 1 + \frac{8x^2y^2}{(x^2 + y^2)^2} \right\}$$

does not exist; since, by allowing $x$ to approach zero first and then $y$ to approach zero, we obtain $-1$, while if the order of the limits is reversed we get $+1$.

The following theorem often gives a convenient test for the interchange of the order of differentiation.

THEOREM III. *If $f''_{xy}$ and $f''_{yx}$ both exist in a region $R$ and each is continuous in $x$ alone and in $y$ alone, then the two are equal at every point of $R$.*

From the continuity of $f''_{xy}$, $f''_{yx}$ in $x$ and in $y$ it follows that the points of $(x, y)$ continuity of these derivative functions form a set $E$ everywhere dense in the given region $R$. As we have seen, these points of continuity are the same for the two derivatives, and at every such point they are equal. Moreover, such points are everywhere dense on any line parallel to either axis of coördinates; for, any line $y = y_0$ may be used for the limiting boundary function

$f(x, y_0)$ and the theorem for the continuity of a function along this line can be applied. Let $(x_0, y_0)$ be any point of $R$ not belonging to $E$. We shall show that $f''_{xy}$ and $f''_{yx}$ are also equal to each other at $(x_0, y_0)$. From the continuity of $f''_{xy}$ with respect to $x$, we have

$$\underset{x \to x_0}{L} f''_{xy}(x, y_0) = f''_{xy}(x_0, y_0).$$

Similarly, since $f''_{yx}$ exists and is continuous in $x$, we have

$$\underset{x \to x_0}{L} f''_{yx}(x, y_0) = f''_{yx}(x_0, y_0).$$

As each of these limits exists, we may obtain the limiting value by allowing $x$ to approach $x_0$ through any set of points dense at $x_0$. Let $\{x_n\}$ be a subset of $E$ on the line

$$y = y_0$$

dense at $x_0$; at each point of which, $f''_{xy}$ and $f''_{yx}$ are equal. We have then

$$\underset{x_n \to x_0}{L} f''_{xy}(x_n, y_0) = \underset{x_n \to x_0}{L} f''_{yx}(x_n, y_0),$$

whence

$$f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0).$$

This theorem gives a sufficient but not a necessary condition as the following example shows.

**Ex. 3.** Given $f(x, y) = x^2y^2 \cos \frac{1}{x}$ for all values of $y$, so long as $x \neq 0$, and let $f(x, y) = 0$, for $x = 0$. Consider the interchange of order of differentiation.

For $x \neq 0$, $y \neq 0$, we have

$$f''_{xy} = f''_{yx} = 4xy \cos \frac{1}{x} + 2y \sin \frac{1}{x}.$$

For $(x = 0, y \neq 0)$, $(x = 0, y = 0)$, and $(x \neq 0, y = 0)$ we obtain

$$f''_{xy} = f''_{yx} = 0.$$

Consequently, these second partial derivatives exist and are equal at every point in every neighborhood of the origin, but neither is continuous in $x$ if $y \neq 0$, since the limit

$$\underset{x \to 0}{L} \left\{4xy \cos \frac{1}{x} + 2y \sin \frac{1}{x}\right\}$$

does not exist. It will be noted, however, that both $f''_{xy}$ and $f''_{yx}$ are continuous in $(x, y)$ together at the origin, since the simultaneous limit exists as $x$ and $y$ approach zero; that is, we have

$$\underset{\substack{x \to 0 \\ y \to 0}}{L} \left\{4xy \cos \frac{1}{x} + 2y \sin \frac{1}{x}\right\} = 0 = f''_{xy}\,(0, 0) = f_{yx}\,(0, 0).$$

The conditions of Theorem I are therefore satisfied.

## EXERCISES

**1.** Given $f(x) = x \sin^2 \frac{1}{x}$, where $f(0) = 0$. Does this function have a first derivative at $x = 0$? Does the second derivative exist at the origin? Show that the limit

$$\underset{\Delta x \to 0}{L} \frac{f(x_0 + \Delta x) - 2f(x_0) + f(x_0 - \Delta x)}{(\Delta x)^2}, \text{ where } x_0 = 0,$$

exists, and explain why its existence does not insure the existence of $f''(0)$.

**2.** Given $f(x) = \sin x \sin \frac{1}{\sin x}$, where $0 \leqq x \leqq 2\pi$. Show that $f'(x)$ does not exist at the points $x = 0,\ \pi,\ 2\pi$.

**3.** Show that the function $f(x) = 4 + 7x + x^2\left(8 + x \sin \frac{1}{x}\right)$, where $x \sin \frac{1}{x}$ is zero for $x = 0$, has a first derivative but no second derivative at the origin.

**4.** Given a function $f(x)$ whose derivative exists everywhere and is finite. Can $f'(x)$ be discontinuous at a set of points of the second category? at the set of rational points? at the irrational points? Can a function $F(x)$ exist whose derivative function $F'(x)$ is continuous where $f'(x)$ is discontinuous and discontinuous at the points of continuity of $f'(x)$?

**5.** By the method of Art. 30 construct the function $y = f(x;\ 3, -2, 3)$ defined as follows and show that it is continuous but can not have a finite derivative at any point. To determine the values of $y$, divide the interval $0 \leqq x \leqq 1$ into three equal parts and the interval $0 \leqq y \leqq 1$ into four equal parts, corresponding to $x = 0, \frac{1}{3}, \frac{2}{3}, 1$, let $y = 0, \frac{3}{4}, \frac{1}{4}, 1$; that is, by the first division $\Delta x = \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ and $\Delta y = \frac{3}{4}, -\frac{2}{4}, \frac{3}{4}$. Continue this method of division indefinitely, thus completely defining the given function for the entire interval. [See *Math. Annalen*, Vol. 52, p. 67.]

**6.** Does the function $f(x) = \frac{1}{\log x^2}$, where $f(0) = 0$, have a derivative for $x = 0$?

**7.** Set up a function (*a*) whose four derivative numbers are distinct at $x = 0$, (*b*) whose right-hand derivative numbers at $x = 0$ are both negative and left-hand both positive, (*c*) one of whose right-hand derivative numbers at $x = 0$ is positive and the other negative, while on the left both are positive.

**8.** Given a function $f(x)$ defined as follows. For $x > 0$, $f(x) = x^{\frac{1}{3}} + a$; for $x < 0$, $f(x) = x^{\frac{1}{3}} - a$; for $x = 0$, $f(x) = 0$. Is $f(x)$ continuous at $x = 0$, and does it have a derivative at that point?

**9.** If $f(x)$ is continuous and has a finite derivative at every point in the closed interval $(0, 1)$, can $f'(x)$ have a removable discontinuity? Can it have a discontinuity of the second kind? Illustrate your conclusion in each case.

**10.** Given $f(x, y) = \frac{x^2 y}{x^4 + y^2}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$. Show that $f_x'$, $f_y$ exist at $(0, 0)$ and examine the continuity of $f_x'$, $f_y'$ with respect to $x$ and to $y$ and $(x, y)$ together.

**11.** Given $f(x, y) = \frac{xy}{x^3 + y^3}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$. Show that $f(x, y)$ is not totally differentiable at $(0, 0)$.

**12.** Let $f(x, y) = \frac{x^2 y^2}{x^4 + y^4}$ for $x \neq 0$, $y \neq 0$, and let $f(0, 0) = 0$. Show that $f_x'$, $f_y'$ are both continuous in $x$ and in $y$, but that $f(x, y)$ is not totally differentiable at the origin. Is $f^{xy}(0, 0) = f_{yx}(0, 0)$?

**13.** Show that it is possible for a function $f(x)$ to be discontinuous at a set of points everywhere dense and still have a finite derivative at a set of points everywhere dense (see *Math. Annalen*, Vol. 70, p. 561).

**14.** Examine the following functions for total differentiability and for interchange of the order of differentiation at the origin.

(*a*) $f(x, y) = \dfrac{xy}{\sqrt{x^2 + y^2}}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$.

(*b*) $f(x, y) = (x^2 + y^2) \sin \dfrac{1}{\sqrt{x^2 + y^2}}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$.

(*c*) $f(x, y) = xy \cdot \dfrac{x^2 - y^2}{\sqrt{x^2 + y^2}}$, $x = 0$, $y \neq 0$, where $f(0, 0) = 0$.

(*d*) $f(x, y) = \log (x^2 + y^4)^{\frac{y^2}{2}}$, $x \neq 0$, $y \neq 0$, where $f(0, 0) = 0$.

(*e*) $f(x, y) = \sqrt{x^2 + y^2} \sin 2\phi$, where $f(0, 0) = 0$, $\phi = \text{arc tan } \dfrac{y}{x}$.

(*f*) $f(x, y) = y^2 \sin \dfrac{x}{y}$, where $f(x, 0) = 0$.

(*g*) $f(x, y) = x^2 \text{ arc tan } \dfrac{y}{x} - y^2 \text{ arc tan } \dfrac{x}{y}$, $x \neq 0$, $y \neq 0$,

where $f(0, 0) = 0$ and $-\dfrac{\pi}{2} < \text{arc tan} < \dfrac{\pi}{2}$.

(*h*) $f(x, y) = | x^2 - y^2 |$.

**15.** Can the partial derivatives $f_x'$, $f_y'$ be discontinuous in $(x, y)$ together and yet $f''_{xy} = f''_{yx}$? Illustrate.

**16.** Given a region $R$ where $f(x, y)$ is continuous in $x$ and $y$ together and $f_x' = 0$, $f_y' = 0$ for all values of $x$ and of $y$ in $R$; show that $f(x, y)$ is a constant in $R$.

**17.** Show that for the function $x^2 \sin \dfrac{1}{x^{\frac{4}{3}}}$, where $f(0) = 0$, the derivative numbers in the neighborhood of $x = 0$ are not bounded, but that the function at the origin has zero as a derivative.

**18.** Show that if $f_x'$, $f_y'$ each have differentials of the first order at a point $(x_0, y_0)$, then $f''_{yx}(x_0, y_0) = f''_{xy}(x_0, y_0)$ and this common value is the double limit

$$\underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{L} \left\{ \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0 + \Delta x, y_0) - f(x_0, y_0 + \Delta y) + f(x_0, y_0)}{\Delta x \, \Delta y} \right\}$$

[See W. H. Young, *Fundamental Theorems in Diff. Cal.*, p. 22.]

**19.** A given function $f(x)$ has a derivative at all points of the interval $(a, b)$. Can $f'(x)$ be infinite at all the rational points? at all the irrational points?

**20.** Could a single-valued function $f(x)$ exist having a derivative at every point of a given interval $(a, b)$, where $f'(x)$ takes at least once the same two distinct values in every subinterval of $(a, b)$?

**21.** Knowing that $f(x)$ has a derivative at every point in an interval $(a, b)$, is one justified without examining the continuity of $f'(x)$ in attempting to find its value at a particular point $x_0$, $a < x_0 < b$, by taking the limit $\underset{x \to x_0}{L} f'(x)$? Illustrate your conclusion.

## CHAPTER V

## RIEMANN THEORY OF INTEGRATION

**36. Definition of a Riemann integral.** From a study of the calculus, the reader has become familiar with the elementary properties of an integral. An integral was considered from two points of view. First of all, integration was considered as the inverse operation of differentiation, and later an integral was regarded as the limit of the sum of an indefinitely large number of small elements. The first notion was employed in evaluating an integral, the latter was often convenient in setting up an integral from given data. By means of the fundamental theorem of the integral calculus, it was seen that these two ways of viewing integration were, for certain classes of functions, only different ways of considering the same problem. It is the purpose of the present discussion to formulate more precisely the definition of an integral and to inquire into the conditions under which an integral may exist.

While some notion of integration, based upon the method of exhaustion, was known to the Greeks, a rigorous treatment of the subject dates from the time of Cauchy and Riemann. The definition of Cauchy covered the case for continuous functions. This definition was extended by Riemann to bounded functions, and he also set up the condition for the existence of such an integral. Later the definition has been further extended to make it applicable to unbounded functions and to integration over unbounded intervals. It is the Cauchy-Riemann definition which is commonly employed in elementary analysis and in the applications to the physical sciences. More recently, other definitions of an integral have been introduced by Lebesgue, Stieltjes, Young and others, which are of special importance in scientific discussions, since they admit a larger range of integrable functions. In this chapter we shall consider the properties of the Riemann integrals, reserving to the following chapter a consideration of the Lebesgue and other integrals.

Let $f(x)$ be a bounded function, defined for the interval $(a, b)$.

Suppose this interval to be subdivided by the insertion between $a(= x_0)$ and $b(= x_n)$ of the intermediate points

$$\sigma \equiv x_1, x_2, \cdot \cdot \cdot , x_{n-1}.$$

Form the sum

$$S(\sigma) = (x_1 - x_0)f(\xi_1) + (x_2 - x_1)f(\xi_2) + \cdot \cdot \cdot + (x_n - x_{n-1})f(\xi_n)$$
$$\equiv \sum_{k=1}^{n} f(\xi_k)\, \Delta_k x,$$

where $\xi_k$ is any point in the interval $(x_k - x_{k-1}) \equiv \Delta_k x$. We now define the Riemann integral as the limit

$$\underset{\Delta \to 0}{L} \sum_{k=1}^{n} f(\xi_k)\, \Delta_k x, \tag{1}$$

provided the value of this limit is independent of the manner of inserting the intermediate points $\sigma$, and $\Delta$ is the largest of the $\Delta_k x$'s, frequently called the **norm** of the given set of intervals. As $\Delta$ approaches zero, $n$ becomes infinite. Symbolically, the Riemann integral is represented by

$$\int_a^b f(x)\, dx.$$

In passing to the limit, it is to be noted that the number of points inserted in each subinterval increases indefinitely as the norm $\Delta$ approaches zero.

It will be seen that the foregoing definition is equivalent to saying that for every arbitrarily small positive number $\epsilon$ there exists a positive number $\delta$, depending on $\epsilon$, such that for every choice of $\xi_k$ in the interval $\Delta_k x$ and for every subdivision whose norm satisfies the condition $\Delta < \delta$, we have

$$\left| \int_a^b f(x)\delta x - \Sigma f(\xi_k)\, \Delta_k x \right| < \epsilon.$$

The function $f(x)$ as defined is bounded in the given interval and the limits of integration are both finite. Integrals which arise under these conditions are called **finite or proper integrals.** Later, we shall see how the notion of integration can be extended to cases where either the function itself becomes infinite at certain points or one or both of the limits of integration are infinite. To distinguish such integrals from those already defined, they are called **infinite or improper integrals.**

**37. Upper and lower integrals.**[1] Denote by $L_k$ and $l_k$ respectively the least upper bound and the greatest lower bound of $f(x)$ in the interval $\Delta_k x \equiv (x_k - x_{k-1})$. Form the two sums

$$\overline{S}(\sigma) = \Sigma\, L_k\, \Delta_k x,\ \underline{S}(\sigma) = \Sigma\, l_k\, \Delta_k x.$$

The values of $\overline{S}(\sigma)$, $\underline{S}(\sigma)$ depend upon the manner in which the given interval $(a, b)$ is subdivided by the insertion of the intermediate points $\sigma$. However, for every method of subdivision, we have

$$\overline{S}(\sigma) \geqq l(b-a),\ \underline{S}(\sigma) \leqq L(b-a), \tag{2}$$

where $L$, $l$ are respectively the least upper bound and the greatest lower bound of $f(x)$ in $(a, b)$. The aggregate of values which $\overline{S}(\sigma)$ may have by all possible methods of subdivision of $(a, b)$ by the insertion of intermediate points has a greatest lower bound which we call the **upper integral** of $f(x)$ in the interval $(a, b)$. Likewise, the sum $\underline{S}(\sigma)$ has for all possible subdivisions of $(a, b)$ a least upper bound which is called the **lower integral** of $f(x)$ in $(a, b)$. We shall denote these integrals respectively by

$$\overline{\int_a^b} f(x)dx,\ \underline{\int_a^b} f(x)dx.$$

We shall now establish the following theorem.

THEOREM. *If $f(x)$ is bounded in the interval $(a, b)$, then*

$$\underset{\Delta\to 0}{L}\ \overline{S}(\sigma) = \overline{\int_a^b} f(x)dx,\quad \underset{\Delta\to 0}{L}\ \underline{S}(\sigma) = \underline{\int_a^b} f(x)dx.$$

We shall prove the theorem for the upper integral. A similar proof establishes the existence of the lower integral and is left to the reader to supply.

By different ways of inserting the intermediate points $\sigma$ in the given interval $(a, b)$ we have different sets of subintervals $\Delta_k$ and may obtain different sums $\overline{S}(\sigma)$. As we have seen, the aggregate of values $\{\overline{S}(\sigma)\}$ has he upper integral $\overline{\int_a^b} f(x)dx$ for its greatest lower bound. We shall now show that

$$\underset{\Delta\to 0}{L}\ \overline{S}(\sigma) = \overline{\int_a^b} f(x)dx, \tag{3}$$

---

[1] The upper and lower integrals were first introduced and rigorously defined by Darboux; see *Ann. Ecol. Normale*, (2), Vol. 4 (1875), p. 64. For that reason they are sometimes called the Darboux integrals.

$\Delta$ being the norm of the $\Delta_k$'s for any method of subdivision whatsoever.

Since $\overline{\int_a^b} f(x)dx$ is the greatest lower bound of $\overline{S}(\sigma)$ for all possible methods of subdivision of the given interval, it follows that we can find a particular method of subdivision for which the sum

$$\overline{S}(\sigma_0) = \sum_{k=1}^{m} L_k\,\delta_k$$

satisfies the relation $\overline{S}(\sigma_0) < \overline{\int_a^b} f(x)dx + \frac{\epsilon}{2}$, (4)

where $\epsilon$ is an arbitrarily small positive number and $\delta_k$ is the length of the subinterval by this method of subdivision.

Consider any other method of subdivision. We can so select the points $\sigma$, and hence $\Delta_k \equiv x_k - x_{k-1}$, that for this new method of subdivision the norm $\Delta$ shall be small enough to satisfy the inequality

$$(m+1)L\cdot\Delta < \frac{\epsilon}{2}, \tag{5}$$

where as before $L$ is the least upper bound of $f(x)$ in $(a, b)$. Some of the resulting intervals $\Delta_k$ may lie wholly within a $\delta_k$ interval while others may contain portions of two or more of the $\delta_k$ intervals. Those $\Delta_k$ intervals falling in the first class contain no end-points of a $\delta_k$ interval, while those of the second class each contain at least one end-point of a $\delta_k$ interval. There can not be more than $(m+1)$ of the $\Delta_k$ intervals of the second class and from (5) these contribute less than $\frac{\epsilon}{2}$ to the sum $\overline{S}(\sigma_0)$. For each of the $\Delta_k$ intervals of the first class we obtain a product $L_k\,\Delta_k$, which is less than the corresponding product that enters into the sum $\overline{S}(\sigma_0)$. Consequently, from (4) we have for all methods of subdivision of $(a, b)$ and for every arbitrarily small value $\epsilon$

$$\overline{S}(\sigma) \leqq \overline{\int_a^b} f(x)dx.$$

But $\overline{\int_a^b} f(x)dx$ is the greatest lower bound of $\overline{S}(\sigma)$, and hence this relation is equivalent to saying that

$$\underset{\Delta\to 0}{L}\ \overline{S}(\sigma) = \overline{\int_a^b} f(x)dx.$$

Since this limiting value is independent of the method by which the given interval is subdivided by the insertion of intermediate points, the theorem follows.

**38. Functions of limited variation.** A class of functions which play an important rôle in the theory of integration is that known as functions of limited variation. We may define this property as follows. Let $f(x)$ be a bounded function defined for the interval $(a, b)$. Suppose this interval to be subdivided by the insertion of the intermediate points

$$x_0(=a), x_1, x_2, x_3, \cdot \cdot \cdot, x_k, \cdot \cdot \cdot x_n(=b).$$

Set up the sum

$$V = |f(x_1) - f(x_0)| + |f(x_2) - f(x_1)| + \cdot \cdot \cdot + |f(x_k) - f(x_{k-1})| \\ + \cdot \cdot \cdot + |f(x_n) - f(x_{n-1})|. \quad (1)$$

The value of $V$ is called the **variation** of $f(x)$ for this particular method of subdivision. Consider the aggregate of values of $V$ obtained by all possible methods of selecting the intermediate points $x_k$. If this set of values is bounded, the function $f(x)$ is said to be of **limited variation.** The least upper bound of the set of values $(V)$ is called the **total variation** of $f(x)$ in the interval $(a, b)$ and is designated by $V_{ab}\,f(x)$.

If in (1) we replace $|f(x_k) - f(x_{k-1})|$ by the oscillation $\omega(f, x_{k-1}x_k)$ of $f(x)$ in the interval $(x_{k-1}, x_k)$, we may write for the total oscillation in $(a, b)$ by this particular subdivision

$$\omega[f(x)] = \omega(f, x_1x_0) + \omega(f, x_2x_1) + \cdot \cdot \cdot + \omega(f, x_nx_{n-1}). \quad (2)$$

If for all methods of subdivision the aggregate of values $(\omega[f(x)])$ has a finite upper bound, then we say that $f(x)$ is of total limited oscillation or simply of **limited oscillation**[1] in the given interval $(a, b)$. Functions which are of limited oscillation are also of limited variation, as the two definitions may be shown to be equivalent.[2] Consequently, either property may be employed as best suits the particular discussion.

If we denote by $p$ the sum of those differences $(f(x_k) - f(x_{k-1}))$ which are positive and by $n$ the sum of the absolute values of those differences which are negative, then from (1) we have

$$V = p - n.$$

---

[1] Sometimes also called "bounded fluctuation," see Hobson, *Theory of Functions of a Real Variable*, 2d. Ed., Vol. I, p. 311.

[2] See Study, *Math. Annalen*, Vol. 47 (1896), p. 298 *et seq.*; see also *Encyclop. des Sci. Math.*, II$_1$, Art. 19, p. 86.

Corresponding to the least upper bound of $V$ in $(a, b)$ there exist least upper bounds $P$ and $N$, of $p$ and $n$, respectively. Moreover, we have

$$f(b) - f(a) = p - n ,$$

whence

$$p = \tfrac{1}{2}\{V + f(b) - f(a)\} ,$$

$$n = \tfrac{1}{2}\{V - f(b) + f(a)\} .$$

Consequently, as $V$ approaches its least upper bound $V_{ab}\,f(x)$, $p$ and $n$ approach respectively their least upper bounds $P$, $N$ for the same subdivisions. The total variation of $f(x)$ in the given interval must then satisfy the relation

$$V_{ab}\,f(x) = P_{ab}\,f(x) + N_{ab}\,f(x). \tag{3}$$

The value of $P$ is called the **total positive variation,** while that of $N$ is called the **total negative variation.** From the foregoing discussion we have also the relation

$$f(b) - f(a) = P_{ab}\,f(x) - N_{ab}\,f(x). \tag{4}$$

We may now establish the following theorems concerning functions of limited variation.

THEOREM I. *If $f(x)$ is of limited variation in a given interval $(a, b)$, then it is also of limited variation in $(a, x')$, where $a < x' < b$.*[1]

For any set of intermediate points between $a$ and $x'$, set up the sum $V'$ for the interval $(a, x')$ similar to (1) for $(a, b)$. The least upper bound of such values $V'$ for all possible sets of intermediate points can not exceed $V_{ab}\,f(x)$. As this latter value is by hypothesis finite, it follows that the least upper bound of $V'$ is finite. Consequently, $f(x)$ is of limited variation in $(a, x')$.

THEOREM II. *A necessary and sufficient condition that a function is of limited variation is that it can be expressed as the difference of two bounded, positive, monotone functions, either both non-increasing or both non-decreasing.*

First of all, the given condition is necessary. We assume $f(x)$ to be of limited variation. If $x$ is any point of the interval $(a, b)$, it follows from Theorem I that $P_{ax}$, $N_{ax}$ are bounded, positive functions of $x$. From (4) we then have

---

[1] In an interval $(a, x)$, the total variation of a function is a function of $x$, say $T(x)$. If the given function $f(x)$ is absolutely continuous in $x$, then $T(x)$ may be shown to be also absolutely continuous in $x$. See Vitali, *Atti. della R. Accad. delle Scienze di Torino,* Vol. 40, p. 1023.

$$f(x) - f(a) = P_{ax}\,f(x) - N_{ax}\,f(x),$$

or

$$f(x) = \{P_{ax}\,f(x) + f(a)\} - N_{ax}\,f(x), \tag{5}$$

from which it follows that $f(x)$ can be expressed as the difference of the two bounded, positive, monotone, non-decreasing functions given in the right-hand member of the above equation. Denote by $A$ any finite number larger than $V_{ab}\,f(x)$. From (5) we have

$$f(x) = \{A - N_{ax}\,f(x)\} - \{A - P_{ax}\,f(x) - f(a)\},$$

which shows that under the conditions set forth in the theorem $f(x)$ may be expressed also as the difference of two bounded, positive, monotone, non-increasing functions.

The given condition is also sufficient. For, suppose that $f(x)$ can be expressed as the difference of the two bounded, non-decreasing, positive functions $F_1(x)$, $F_2(x)$. We have then for any subinterval $(x_{k-1}, x_k)$ of $(a, b)$

$$|f(x_k) - f(x_{k-1})| \leqq \{F_1(x_k) - F_1(x_{k-1})\} + \{F_2(x_k) - F_2(x_{k-1})\}.$$

Taking the sum of these values for all subintervals of $(a, b)$, we get

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leqq \{F_1(b) - F_1(a)\} + \{F_2(b) - F_2(a)\}.$$

This relation holds for all methods of subdivision, and as the second member of the inequality is finite, it follows from the definition of limited variation that $f(x)$ has that property in the given interval $(a, b)$.

A similar proof holds for the case where $F_1(x)$, $F_2(x)$ are both bounded, positive, monotone, non-increasing functions.

It may also be shown that any absolutely continuous function may be expressed as the difference of two non-decreasing, absolutely continuous functions.[1]

THEOREM III. *If two given functions are each of limited variation in $(a, b)$, then their sum, difference, and product are also of limited variation in the same interval.*

Denote the given functions by $f_1(x)$ and $f_2(x)$. From Theorem II, each can be expressed as the difference of two bounded, monotone, positive, non-decreasing or non-increasing functions. We shall consider the case where $f_1(x)$, $f_2(x)$ are the difference of two non-decreasing functions. For the other case the proof is similar. We have then

---

[1] See Vitali, *Atti della R. Accad. delle Scienze di Torino*, Vol. 40, p. 1024.

$$f_1(x) = \phi_1(x) - \psi_1(x), \; f_2(x) = \phi_2(x) - \psi_2(x),$$

where $\phi_1$, $\phi_2$, $\psi_1$, $\psi_2$ are in $(a, b)$ bounded, monotone, positive, non-decreasing functions. The sum, difference, and product of $f_1(x)$, $f_2(x)$ can now be expressed as the difference of two monotone, positive, non-decreasing functions as follows.

$$\begin{aligned} f_1(x) + f_2(x) &= \{\phi_1 + \phi_2\} - \{\psi_1 + \psi_2\}, \\ f_1(x) - f_2(x) &= \{\phi_1 + \psi_2\} - \{\phi_2 + \psi_1\}, \\ f_1(x) \cdot f_2(x) &= \{\phi_1\phi_2 + \psi_1\psi_2\} - \{\phi_2\psi_1 + \phi_1\psi_2\}. \end{aligned}$$

Since the sum, the difference, and the product can be expressed as the difference of two bounded, monotone, positive, non-decreasing functions, it follows from Theorem II that each is a function of limited variation.

THEOREM IV. *If $f(x)$ is of limited variation, then $\frac{1}{f(x)}$ is also, provided $|f(x)| \geqq M$, where $M$ is a positive number.*

From (1) we have for $\frac{1}{f(x)}$

$$\begin{aligned} V = \sum_{k=1}^{n} \left| \frac{1}{f(x_k)} - \frac{1}{f(x_{k-1})} \right| &= \sum_{k=1}^{n} \left| \frac{f(x_{k-1}) - f(x_k)}{f(x_k)f(x_{k-1})} \right| \\ &\leqq \frac{1}{M^2} \sum_{k=1}^{n} \left| f(x_k) - f(x_{k-1}) \right|. \end{aligned}$$

But as $f(x)$ is of limited variation, the sum $\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|$ is bounded for all methods of choosing the intermediate points $x_k$. Consequently, the value of $V$ for $\frac{1}{f(x)}$ is, under the conditions set forth in the theorem, bounded and $\frac{1}{f(x)}$ is therefore of limited variation, as the theorem requires.

THEOREM V. *A function of limited variation can have only ordinary discontinuities.*

We shall assume that at some point, say $x_0$, of the given interval $(a, b)$, $f(x)$ has discontinuity of the second kind and show that under this assumption $f(x)$ can not be of limited variation in $(a, b)$. The right-hand and left-hand limits

$$\underset{x \to x_0+}{L} f(x), \quad \underset{x \to x_0-}{L} f(x)$$

do not exist, since $x_0$ is a point of discontinuity of the second kind. However, there exist at $x_0$ a maximum and a minimum value of $f(x)$ on both the right-hand and the left-hand. We have then

$$\overline{f(x_0 + 0)} > \underline{f(x_0 + 0)},$$

where $\overline{f(x_0 + 0)} \equiv \underset{\epsilon \to 0}{L} \overline{f(x_0 + \epsilon)}, \quad \underline{f(x_0 + 0)} \equiv \underset{\epsilon \to 0}{L} \underline{f(x_0 + \epsilon)}.$

It follows that to the right of $x_0$ it is possible to find a set of points

$$\{x_k\} \equiv x_1, x_2, \cdot \cdot \cdot , x_k, \cdot \cdot \cdot$$

dense at $x_0$ such that we have

$$| f(x_k) - f(x_{k-1}) | > A,\ k = 1, 2, \cdot \cdot \cdot ,$$

where $A$ is a suitably chosen positive number. As the points $x_k$ are dense at $x_0$ the number $n$ of the intervals $(x_k - x_{k-1})$ in (1) lying within $(a, b)$ can be made as large as one may choose. From (1) we have for the foregoing subdivision of $(a, b)$

$$V = \sum_{k=1}^{n} | f(x_k) - f(x_{k-1}) | > nA.$$

As $A$ is by hypothesis different from zero and $n$ can be taken as large as we choose, it follows that $V$ is not bounded and consequently $f(x)$ can not be of limited variation in $(a, b)$. From this conclusion the truth of the theorem follows.

THEOREM VI. *The points of discontinuity of a function of limited variation form at most an enumerable set.*[1]

This theorem follows as a consequence of Theorem VI, Art. 28, when we remember that a function of limited variation can have only ordinary discontinuities.

A relation between absolute continuity and the limited variation of a function is expressed by the following theorem.

THEOREM VII. *If $f(x)$ is absolutely continuous in an interval $(a, b)$, then it is also of limited variation in this interval.*

---

[1] This theorem also follows from the fact that a function of limited variation can be expressed as the difference of two monotone functions and every monotone function can have at most an enumerable set of points of discontinuity. See Carathéodory, *Vorlesungen*, p. 154.

It may be shown that the points of discontinuity of a function of limited variation form a set of the first species. See Lebesgue's *Leçons sur l'intégration*, p. 58.

For, if $f(x)$ is not of limited variation, then there must be at least one subinterval of $(a, b)$ in which the total variation of $f(x)$ exceeds every finite bound, however small the interval is taken. In this interval, the sum of the oscillations of $f(x)$ must also exceed every finite bound. It follows that in any set of subintervals including one of this character the sum of the oscillations of $f(x)$ can not tend toward zero with the sum of the intervals. Consequently, $f(x)$ can not be said to be absolutely continuous in $(a, b)$. From this contradiction to the given hypothesis the theorem follows.

The converse of this theorem is not true; that is, there are functions of limited variation which are not absolutely continuous. As a matter of fact, a function of limited variation may have an enumerably infinite number of points of ordinary discontinuity and hence is not even continuous, to say nothing of being absolutely continuous. A function may be both continuous and of limited variation and still not be absolutely continuous.[1] It may be shown, however, that a continuous function of limited variation, whose derivative numbers are infinite at most over an enumerable set $E$, is absolutely continuous.[2]

Not every continuous function is of limited variation. For example, suppose we consider

$$\begin{aligned} f(x) &= x \sin \frac{1}{x}, \text{ for } x \neq 0, \\ &= 0, \text{ for } x = 0. \end{aligned}$$

As we know, this function is continuous in the closed interval (0, 1). It has an infinite number of oscillations in every neighborhood of the origin, there being one in every subinterval $\left(\frac{1}{k\pi}, \frac{1}{(k+2)\pi}\right)$. The amount of the oscillation in any one of these subintervals is greater than $\frac{2}{(k+2)\pi}$. Hence, the variation $V_{01}\,[f(x)]$ satisfies the condition

$$V_{01}\,[f(x)] > \Sigma \frac{2}{(k+2)\pi} = \frac{2}{\pi}\left\{\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots + \frac{1}{n} + \cdots\right\}.$$

However, the infinite series in the brackets is divergent, and hence the variation is not limited.

The foregoing function has no derivative at the origin. Even a function having a finite derivative at every point may nevertheless be of unlimited variation. The function,

---

[1] See Vitali, *Atti della R. Accad. delle Scienze di Torino* (1905), Vol. 40, p. 1022.

[2] See Porter, *Bull. Amer. Math. Soc.* (1915), Ser. 2, Vol. 22, p. 110.

$$\begin{aligned} f(x) &= x^2 \sin \frac{1}{x^2}, \text{ for } x \neq 0, \\ &= 0, \text{ for } x = 0, \end{aligned}$$

is an illustration. This function has a finite derivative at every point in the interval (0, 1). One may show that it is of unlimited variation in (0, 1) by following the method of the previous illustration.

It is of interest to note that $f(x)$ as defined in either of these illustrations is not absolutely continuous. For, if it were absolutely continuous, it would be of limited variation by the foregoing theorem, and as we have seen, that is not the case. One must not infer that simply because a function has an infinite number of oscillations in a finite interval, it must be of unlimited variation. For, let

$$\begin{aligned} f(x) &= x^2 \sin \frac{1}{x^{2t}}, \text{ for } t < 1,\ x \neq 0, \\ &= 0, \text{ for } x = 0. \end{aligned}$$

This function vanishes for $x = \dfrac{1}{(k\pi)^{\frac{1}{2t}}}$ and has an oscillation in every interval $\left(\dfrac{1}{(k\pi)^{\frac{1}{2t}}}, \dfrac{1}{[(k+2)\pi]^{\frac{1}{2t}}}\right)$. The amount of the oscillation is less than

$$2x^2 = \frac{2}{(k\pi)^{\frac{1}{t}}}.$$

Hence, the sum of the oscillations in any neighborhood of the origin is less than

$$\frac{2}{\pi^{\frac{1}{t}}}\left\{1 + \frac{1}{2^{\frac{1}{t}}} + \frac{1}{3^{\frac{1}{t}}} + \cdots + \frac{1}{n^{\frac{1}{t}}} + \cdots\right\},$$

which is convergent for any value of $t$ less than one. The function is therefore of limited variation, since by any other method of selecting the intervals the sum of the oscillations could not be increased.

THEOREM VIII. *If any one of the derivative numbers of a function $f(x)$ is bounded for all values of $x$ in the interval $(a, b)$, then $f(x)$ is continuous and of limited variation.*

If one of the derivative numbers is bounded, then all are bounded and have the same least upper and greatest lower bounds as the difference quotient. We may write for every pair of values $x_1$, $x_2$ in $(a, b)$

$$\left|\frac{f(x_1) - f(x_2)}{x_1 - x_2}\right| < M,$$

where $M$ is a finite number. The total variation of $f(x)$ in $(a, b)$ can not exceed $2M(b - a)$. Since this product is finite, it follows that $f(x)$ is of limited variation in the given interval.

The function $f(x)$ is continuous; for, a function of limited variation can have only ordinary discontinuities (Theorem V) and this can not occur, since in that case at least some of the derivative numbers are no longer bounded.

The converse of the foregoing theorem is not true; that is, there are continuous functions of limited variation which do not have bounded derivative numbers for some values of $x$. For example, the function

$$f(x) = \sqrt{|x|}$$

is continuous and of bounded variation in any interval including the origin, but its derivative numbers at the origin are not bounded.

**39. Existence theorems for Riemann integrals.** We shall consider the conditions under which a Riemann integral may exist; that is, the conditions which must be imposed upon the function $f(x)$ in order that the limit (1) of Art. 36 may exist. The first of these theorems may be stated as follows.

THEOREM I. *Given a function $f(x)$ which is bounded in the interval $(a, b)$. A necessary and sufficient condition that the integral $\int_a^b f(x)dx$ exists is that*

$$\overline{\int}^b f(x)dx = \underline{\int_a}^b f(x)dx.$$

We may show that this condition is sufficient by considering the inequality,

$$S(\sigma) \leqq S(\sigma) \leqq \bar{S}(\sigma), \tag{1}$$

where $\underline{S}(\sigma)$, $S(\sigma)$, $\bar{S}(\sigma)$ have the values assigned to them in Arts. 36, 37. Since by any method of subdivision we have

$$\underset{\Delta \to 0}{L}\,\underline{S}(\sigma) = \underline{\int_a}^b f(x)dx, \quad \underset{\Delta \to 0}{L}\,\bar{S}(\sigma) = \overline{\int_a^b} f(x)dx,$$

and since by hypothesis the upper and the lower integrals are equal, it follows from Theorem IV, Art. 24, that the limit

$$\underset{\Delta \to 0}{L}\,S(\sigma)$$

exists and $f(x)$ is integrable, the integral $\int_a^b f(x)dx$ being the common value of the upper and the lower integrals.

The given condition is also necessary. To show this, we have the condition that the limit

$$\underset{\Delta\to 0}{L}\ \sum_{k=1}^{n} f(\xi_k)\,\Delta_k x$$

exists for all methods of subdivision of the given interval and for all values of $\xi_k$ in $\Delta_k x$. It follows that the limit still exists if $f(\xi_k)$ is replaced by the least upper bound $L_k$ or by the greatest lower bound $l_k$ of $f(x)$ in $\Delta_k x$. However, when these values are substituted, we have nothing else than the upper and the lower integrals $\overline{\int_a^b} f(x)dx$, $\underline{\int_a^b} f(x)dx$, respectively, and their existence and equality follow as a consequence of the fact that they are special cases of the given limit.

It is sometimes convenient to have the condition of integrability expressed in terms of the oscillation of the function. Denote by $\omega_k$ the oscillation of $f(x)$ in the interval $\Delta_k x$; that is, put

$$\omega_k \equiv \omega(f, \Delta_k).$$

We can then form the oscillatory sum

$$\Omega_n[f, ab] \equiv \sum_{k=1}^{n} \omega_k\,\Delta_k x,$$

and state the following theorem.

THEOREM II. *A necessary and sufficient condition that a bounded function $f(x)$ is integrable in the Riemann sense is that*

$$\underset{\Delta\to 0}{L}\ \Omega_n[f, ab] = 0.$$

The sum $\Omega_n$ may be written

$$\begin{aligned}\Omega_n[f, ab] &= \sum_{k=1}^{n} \{L_k - l_k\}\,\Delta_k x\\ &= \bar{S}(\sigma) - \underline{S}(\sigma).\end{aligned}$$

As each of the limits $\underset{\Delta\to 0}{L}\ \bar{S}(\sigma)$, $\underset{\Delta\to 0}{L}\ \underline{S}(\sigma)$ exists, we then have from the condition given in the theorem

$$\underset{\Delta\to 0}{L}\ \Omega_n[f, ab] = \underset{\Delta\to 0}{L}\ \{\bar{S}(\sigma) - \underline{S}(\sigma)\} = 0,$$

and hence

$$\underset{\Delta\to 0}{L}\ \bar{S}(\sigma) = \underset{\Delta\to 0}{L}\ \underline{S}(\sigma).$$

By Theorem I, it follows that this common value gives the integral $\int_a^b f(x)dx$; that is, we have in the limit

$$\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx = \int_a^b f(x)dx.$$

This conclusion shows that the given condition is sufficient.

The given condition is also necessary; for, if the integral exists, we have

$$\overline{\int_a^b} f(x)dx = \underline{\int_a} f(x)dx,$$

or

$$\underset{\Delta\to 0}{L}\, \overline{S}(\sigma) = \underset{\Delta\to 0}{L}\, \underline{S}(\sigma) = A,$$

whence

$$\underset{\Delta\to 0}{L}\, \overline{S}(\sigma) - \underset{\Delta\to 0}{L}\, \underline{S}(\sigma) = \underset{\Delta\to 0}{L}\, \{\overline{S}\ (\sigma) - \underline{S}(\sigma)\} = \underset{\Delta\to 0}{L}\, \Omega_n[f, ab] = 0.$$

We have also the following theorem due to Riemann.[1]

THEOREM III. *Given a function $f(x)$ which is bounded in the interval $(a, b)$. A necessary and sufficient condition that the Riemann integral of $f(x)$ exists in $(a, b)$ is that this interval may be divided into partial intervals such that the sum of the lengths of those subintervals in which the oscillation of $f(x)$ is equal to or greater than any arbitrarily chosen positive number $\eta$ may be made as small as one pleases.*

We may show as follows that the given condition is sufficient. Denote by $S_m'$ the sum of the lengths of the subintervals of $(a, b)$ in which the oscillation of $f(x)$ is equal to or greater than $\eta$, and let $C(S_m')$ be the sum of the lengths of the remaining subintervals of $(a, b)$; that is, the sum of the lengths of those subintervals in which the oscillation is less than $\eta$. We may then write

$$\Omega_n[f, ab] < \eta(b - a) + S_m' \cdot \omega, \tag{2}$$

where $\omega$ is the greatest value of the oscillation $\omega_k$ of $f(x)$ in those subintervals included in $S_m'$. The value of $\omega$ is finite and can not increase as the size of the intervals is decreased by making the norm $\Delta$ to approach zero. By hypothesis, we have

$$S_m' < \epsilon,$$

where $\epsilon$ is positive and may be chosen arbitrarily small. We then have from (2)

$$\Omega_n[f, ab] < \eta(b-a) + \epsilon_1,$$

where $\epsilon_1$ is arbitrarily small. Since this relation holds for every arbitrary choice of $\eta$ and $\epsilon_1$ however small, we have

$$\underset{\Delta\to 0}{L}\, \Omega_n[f, ab) = 0,$$

and consequently $f(x)$ is integrable by Theorem II.

---

[1] See *Habilitationschrift*, Göttingen, 1854, §5; see also *Gesammalte Werke* (2d Ed., 1892), p. 240.

To show that this condition is also necessary, we proceed as follows. We assume that $f(x)$ is integrable in $(a, b)$, and hence we have from Theorem II

$$\underset{\Delta\to 0}{L}\ \Omega_n[f, ab] = 0. \tag{3}$$

However, we may write

$$\Omega_n[f, ab] = \sum_{k=1}^{n} \omega_k \Delta_k x \geqq \eta \cdot S'_m + \omega \cdot C(S'_m) \geqq \eta \cdot S'_m, \tag{4}$$

where $\omega$ is the smallest value of $\omega_k$ for the $n$ subdivisions of $(a, b)$. Making use of (3) we have from (4), by passing to the limit,

$$\underset{\Delta\to 0}{L}\ \Omega_n[f, ab] = 0 \geqq \eta \cdot \underset{\Delta\to 0}{L}\ S'_m$$

While $\eta$ is arbitrarily chosen, it is greater than zero. Since both $\eta$ and $S'_m$ are positive we must have

$$\underset{\Delta\to 0}{L}\ S'_m = 0;$$

that is, we have

$$S'_m < \epsilon,$$

where $\epsilon$ is arbitrarily small.

It follows from this theorem that if the points of discontinuity where the saltus exceeds $\eta$ form a set of the first species, then $f(x)$ is integrable. These points may also form a set of the second species and still have $f(x)$ integrable in the given interval, but this is not necessarily true as is the case where the points in question form a set of the first species.

As a corollary to the foregoing theorem we may state the following:

COR. *A necessary and sufficient condition that a given bounded function is integrable is that its points of discontinuity where the saltus is equal to or greater than an arbitrarily chosen member $\eta$ form a discrete set.*

By introducing the notion of measure, we may state the following important criterion for integration.

THEOREM IV. *A necessary and sufficient condition that a bounded function $f(x)$ is integrable in the Riemann sense is that in the given interval the points of discontinuity form a set of measure zero.*

To show that the condition is necessary, consider a sequence of values

$$\eta_1 > \eta_2 > \eta_3 > \cdot \cdot \cdot > \eta_n > \cdot \cdot \cdot,$$

where

$$\underset{n\to\infty}{L}\ \eta_n = 0,$$

and let $(a, b)$ be the interval within which $f(x)$ is bounded and integrable. Denote by $E_n$ $(n = 1, 2, \cdots)$ the set of points of $(a, b)$ at each of which $f(x)$ has a discontinuity with a saltus equal to or greater than $\eta_n$. The sequence of point sets $\{E_n\}$ satisfies the condition that

$$E_1 < E_2 < \cdots < E_n < \cdots.$$

All of the points of $E_n$ belong to subintervals of $(a, b)$ within which the oscillation of $f(x)$ is at least equal to $\eta_n$. By Theorem III, the sum of these subintervals may be made as small as we please and hence the measure of the set $E_n$ is zero.

All the points of discontinuity $E$ of $f(x)$ in $(a, b)$ are included in the sum

$$E_1 + E_2 + \cdots + E_n + \cdots.$$

The measure of $E$, by Theorem VIII, Art. 21, is

$$\underset{n\to\infty}{L}\, m(E_n) = \underset{n\to\infty}{L}\, 0 = 0.$$

Hence, it follows that the conditions of the theorem are necessary for the integrability of $f(x)$ in $(a, b)$.

The given condition is also sufficient. For, the set $E_0$ of points of discontinuity of $f(x)$ where the saltus is equal to or greater than a given number $\eta$ is a closed subset of $E$. By hypothesis the set $E$ is of measure zero, and hence the subset $E_0$ is also of measure zero. But $E_0$ is closed and consequently its content is identical with its measure. By the corollary to Theorem III, $f(x)$ is therefore integrable.

By aid of these theorems, we are now able to determine certain classes of functions for which the Riemann integral always exists. These results will be stated in the following theorems.

THEOREM V. *If $f(x)$ is continuous in the closed interval $(a, b)$, then the integral $\int_a^b f(x)dx$ exists.*

It follows at once, since $(a, b)$ is a closed interval, that $f(x)$ is bounded and hence the foregoing theorems apply. As there are no points of discontinuity, the conditions of Theorem III are satisfied and the theorem follows.

THEOREM VI. *A bounded function $f(x)$ having a finite or enumerably infinite number of discontinuities is integrable in the Riemann sense.*

As we have seen, every enumerable set is of measure zero. It then follows from Theorem IV of this article that $f(x)$ is integrable.

From this theorem one sees that a bounded function may be discontinuous at the set of rational points and still be integrable, provided it is continuous at the irrational points.

THEOREM VII. *A function $f(x)$ of limited variation in the closed interval $(a, b)$ is integrable in the Riemann sense.*

The given function is bounded and by Theorem VI, Art. 38, its points of discontinuity form at most an enumerable set. Consequently by Theorem VI it follows that $f(x)$ is integrable.

THEOREM VIII. *A bounded function $f(x)$ having only ordinary finite discontinuities in the given interval $(a, b)$ is integrable in the Riemann sense.*

By Theorem VI, Art. 28, it follows that the points where the given function has ordinary discontinuities form an enumerable set. From Theorem VI it follows that $f(x)$ is integrable in the interval for which it is defined.

THEOREM IX. *Every bounded, monotone function $f(x)$ is integrable in the Riemann sense.*

Let the given interval be $(a, b)$ and suppose that $\epsilon$ as usual is an arbitrarily chosen positive number. By taking $n$, the number of divisions of $(a, b)$, sufficiently large, we can make the norm $\Delta$ of the subintervals so small that

$$\Delta = \frac{\epsilon}{f(b) - f(a)}.$$

We have for the oscillatory sum

$$\Omega_n[f, ab] \leqq \Delta \sum_{k=1}^{n} \{f(x_k) - f(x_{k-1})\} = \Delta \cdot \{f(b) - f(a)\} = \epsilon.$$

Passing to the limit, this relation becomes

$$\underset{\Delta \to 0}{L} \ \Omega_n = 0,$$

and the given function is integrable by Theorem II.

Since every bounded, monotone function is integrable, and the points of discontinuity of an integrable function form a set of measure zero, we have at once the following corollary.

COR. *If $f(x)$ is a bounded, monotone function, then its points of discontinuity form at most a set of measure zero.*

It is to be noted that not every point-wise discontinuous function is integrable. While a necessary and sufficient condition for a function to be point-wise discontinuous is that its points of discontinuity where the saltus exceeds a given number $\epsilon$ forms a set in no subinterval everywhere dense and while a discrete set can not be everywhere dense in any portion of the given interval, it is true that there also exist point sets which are not discrete which also satisfy this condition. We can not say, therefore, that a point-wise discontinuous function $f(x)$ is integrable unless it satisfies still other conditions.

**40. Integrals of bounded functions.** We shall limit our discussion to those properties of definite integrals not commonly discussed in an elementary course in calculus, and in the present article the discussion is concerned with bounded functions. Among the more important theorems relating to definite integrals of bounded functions we have the following:

THEOREM I. *If $f(x)$ is integrable in $(a, b)$, it is integrable in any subinterval of $(a, b)$.*

Let $(a_1, b_1)$ be any subinterval of $(a, b)$. By the corollary to Theorem III of the previous article, $f(x)$ can have in $(a, b)$ at most a discrete set $E$ of points of discontinuity where the saltus is equal to or greater than an arbitrarily chosen number $\eta$. It follows at once that the points having the same property and situated in $(a_1, b_1)$ must likewise form a discrete set, since it is a subset of $E$. Consequently, it follows also from the corollary to Theorem III of the last article that $f(x)$ is integrable in $(a_1, b_1)$, as was to be shown.

THEOREM II. *Given a finite number of functions $f_1(x), f_2(x), \cdots, f_n(x)$ each of which is integrable in $(a, b)$. It follows that*

$$\int_a^b \{f_1(x) + f_2(x) + \cdots + f_n(x)\}\, dx = \int_a^b f_1(x)dx + \int_a^b f_2(x)dx + \cdots + \int_a^b f_n(x)dx.$$

This theorem follows at once from the definition of a definite integral as a limit and the fact that each of the $n$ limits exists; for, the limit of the sum is then the sum of the separate limits. This theorem still holds if each function is multiplied by a constant.

The following theorem, due to Du Bois-Reymond, is of importance in the discussion of integrable functions.[1]

THEOREM III. *If* $\varphi_1(x)$, $\varphi_2(x)$, $\cdots$ $\varphi_n(x)$ *are a finite number of bounded functions integrable in the closed interval* $(a, b)$, *then any continuous function* $F(\varphi_1, \varphi_2, \cdots, \varphi_n)$, *considered as a function of* $x$, *is integrable in* $(a, b)$.

We shall prove the theorem for the case of two functions $\varphi_1(x)$, $\varphi_2(x)$. The argument is the same for the case where $n$ has any finite value.

As $x$ takes the values in the interval $(a, b)$, suppose that the values of $\varphi_1$, $\varphi_2$ lie within the closed intervals $(c_1, d_1)$ and $(c_2, d_2)$, respectively. Within the region defined by these intervals, $F(\varphi_1, \varphi_2)$ is by hypothesis a continuous function of $\varphi_1$, $\varphi_2$. Put

$$f(x) \equiv F[\varphi_1(x), \varphi_2(x)].$$

The only points of discontinuity which $f(x)$ can have are those of $\varphi_1(x)$ and $\varphi_2(x)$; for, at any other point $f(x)$ is a continuous function of a continuous function and hence is continuous.

For any arbitrarily selected $\eta_n$ the points of discontinuity of $\varphi_1$, $\varphi_2$ where the saltus is equal to or greater than $\eta_n$ form a closed discrete set and hence their sum, which we shall denote by $E_n$, is also closed and discrete. Let $\eta_n$ take a sequence of values

$$\eta_1 > \eta_2 > \cdots > \eta_n > \cdots,$$

having the limit zero as $n$ becomes infinite. Corresponding to these values, we have a sequence of point sets $\{E_n\}$ and we shall denote the limiting set of this sequence by $E$. The set of points of discontinuity $G$ of $f(x)$ for which the saltus is greater than an arbitrarily small $\eta$ is a closed subset of $E$. By Theorem V of Art. 20, it follows that $G$ is discrete and consequently $f(x)$ is integrable in the interval $(a, b)$.

It follows from the foregoing theorem that any rational function of an integrable function is integrable, provided the denominator does not vanish in the given interval. Thus, the product of a finite number of integrable functions or any integral power of an integrable function is integrable. Likewise the quotient of one integrable function by another integrable function is integrable if the function by which one divides is not zero in the given interval.

---

[1] *Math. Annalen*, Vol. 20 (1882), p. 122

From Theorem II, we saw that the integral of a sum of a finite number of integrable functions is the sum of their integrals. It must not be inferred from the present theorem that the integral of any rational function of integrable functions is that function of the integrals. For example, it does not follow that the integral of the product of two integrable functions is the product of their integrals.

The question naturally suggests itself whether one might not generalize the foregoing theorem by showing that every integrable function of an integrable function is integrable, just as we have seen that every continuous function of a continuous function is continuous. That this generalization is not possible is shown by the following illustration.[1]

Let $y = \phi(x)$ be defined as equal to zero for irrational values of $x$ and equal to $\frac{1}{q}$ for $x$ equal to $\frac{p}{q}$, where $p$ and $q$ are prime to each other; that is, for rational values of $x$. Define $f(y)$ as equal to one, except for $y$ equal to zero, where it takes the value zero. The integral $\int_0^1 \phi(x)dx$ exists by Theorem VI of Art. 39, since $\phi(x)$ is bounded and has at most discontinuities at the rational points; that is, at an enumerable set of points. Moreover, the integral $\int_0^1 f(y)dy$ exists as it has but one discontinuity and that is finite. However, the function

$$F(x) = f[\phi(x)]$$

is not integrable; for, it is totally discontinuous, having the value one at all rational points and zero at all irrational points.

THEOREM IV. *Given two bounded functions having the same points of continuity. If one is integrable, then the other is also.*[2]

Let the two functions be $\phi_1(x)$, $\phi_2(x)$ and suppose that one, say $\phi_1(x)$, is integrable. By Theorem IV of the last article, the points of discontinuity of $\phi_1(x)$ must form a set of measure zero. Since $\phi_2(x)$ has the same points of continuity as $\phi_1(x)$, its points of discontinuity form at most a set of measure not greater than that of the discontinuities of $\phi_1(x)$. It must then have the measure zero and hence $\phi_2(x)$ is integrable, as the theorem states.

THEOREM V. *If $f(x)$ and $\phi(x)$ are both bounded, integrable functions in the interval $(a, b)$ and $f(x)$ is positive, then*

$$F(x) \equiv [f(x)]^{\phi(x)}$$

*is also integrable in $(a, b)$.*

---

[1] See Lebesgue, *Leçons sur l'intégration*, p. 30.

[2] See W. H. Young, *Quarterly Journal*, Vol. 35 (1904), p. 190.

For those values of $x$ for which both $f(x)$ and $\phi(x)$ are continuous, $F(x)$ is also continuous as a consequence of Theorem II, Art. 24, and the definition of continuity. The only points of discontinuity that $F(x)$ can have are those points for which either $f(x)$ or $\phi(x)$ or both, are discontinuous. As $f(x)$ and $\phi(x)$ are both integrable, their points of discontinuity form sets of measure zero and hence by Theorem VI, Art. 21, those of $F(x)$ must also have the measure zero, since the measure of the points of discontinuity of $F(x)$ can not exceed the sum of the two sets of measure zero having no points in common, namely the discontinuities of $f(x)$ and those of $\phi(x)$ not points of discontinuity of $f(x)$. It follows that $F(x)$ is integrable in $(a, b)$.

THEOREM VI. *If $f(x)$ is bounded and integrable in $(a, b)$, then $|f(x)|$ is integrable in the same interval and*

$$\left| \int_a^b f(x)dx \right| \leqq \int_a^b |f(x)|\, dx.$$

We shall first show that the integral in the second member of this inequality exists if the integral $\int_a^b f(x)dx$ exists. By comparison we see that the oscillation of $|f(x)|$ in any interval is less than or at most equal to that of $f(x)$. By Theorem II, Art. 39, we have

$$\underset{\Delta\to 0}{L}\ \Omega_n[f, ab] = 0,$$

where

$$\Omega_n = \sum_{k=1}^{n} \omega_k\, \Delta_k(x).$$

But since the oscillation of $|f(x)|$ can not exceed that of $f(x)$ in any subinterval $\Delta_k(x)$, it follows that the oscillatory sum $\Omega_n'$ formed for $|f(x)|$ must be equal to or less than $\Omega_n$. Consequently, we have

$$\underset{\Delta\to 0}{L}\ \Omega_r' = 0,$$

and the integral $\int_a^b |f(x)|\, dx$ must also exist.

From the definition of a definite integral, we have

$$\left| \int_a^b f(x)dx \right| = \left| \underset{\Delta\to 0}{L}\ \sum_{k=1}^{n} f(\xi_k)\Delta_k(x) \right|,$$

$$\int_a^b |f(x)|\, dx = \underset{\Delta\to 0}{L}\ \sum_{k=1}^{n} |f(\xi_k)|\, \Delta_k(x).$$

But since the absolute value of a sum is less than or at most equal to the sum of the absolute values, we have for all values of $\Delta$, and hence for all values of $n$,

$$\left| \sum_{k=1}^{n} f(\xi_k)\Delta_k(x) \right| \leqq \sum_{k=1}^{n} |f(\xi_k)|\, \Delta_k(x).$$

Passing to the limit as $\Delta$ approaches zero, we have

$$\left| \int_a^b f(x)dx \right| \leqq \int_a^b |f(x)|\, dx,$$

which establishes the theorem.

The converse of this theorem is not true; that is, $|f(x)|$ can be integrable even when $f(x)$ is not integrable. For example, if $f(x)$ is $\frac{1}{2}$ for all rational values of $x$ and $-\frac{1}{2}$ for all irrational values, the integral $\int_0^1 |f(x)|\, dx$ exists although the integral $\int_0^1 f(x)dx$ does not exist, since $|f(x)|$ is continuous while $f(x)$ is totally discontinuous.

THEOREM VII. *If $f(x)$ and $\phi(x)$ are bounded and integrable in the interval $(a, b)$ and if $|f(x)| \leqq |\phi(x)|$ for all values of $x$ in $(a, b)$, it follows that*

$$\left| \int_a^b f(x)dx \right| \leqq \int_a^b |\phi(x)|\, dx.$$

From the condition

$$|f(x)| \leqq |\phi(x)|,$$

it follows that each term entering into the summation defining the integral $\int_a^b |f(x)|\, dx$ can not exceed the corresponding term in the summation defining the integral $\int_a^b |\phi(x)|\, dx$, and hence

$$\int_a^b |f(x)|\, dx \leqq \int_a^b |\phi(x)|\, dx.$$

But by Theorem VI, we have

$$\left| \int_a^b f(x)dx \right| \leqq \int_a^b |f(x)|\, dx.$$

By a comparison of these two inequalities, it follows that

$$\left| \int_a^b f(x)dx \right| \leqq \int_a^b |\phi(x)|\, dx,$$

as the theorem states.

THEOREM VIII. *If $f(x)$ is a bounded, integrable function in the interval $(a, b)$, then*

$$m(b-a) \leqq \int_a^b f(x)\, dx \leqq M(b-a),$$

*where $m$ and $M$ are respectively the greatest lower and the least upper bounds of $f(x)$ in $(a, b)$.*

We have at once from the conditions of the theorem

$$\sum_{k=1}^{n} m\Delta_k x \leqq \sum_{k=1}^{n} f(\xi_k)\Delta_k x \leqq \sum_{k=1}^{n} M\Delta_k x.$$

Passing to the limit as the norm $\Delta x$ approaches zero, we have from the definition of an integral

$$m\int_a^b dx \leqq \int_a^b f(x)dx \leqq M\int_a^b dx,$$

or

$$m(b-a) \leqq \int_a^b f(x)dx \leqq M(b-a),$$

which is the desired result. This result is called the **First Law of the Mean** for the integral calculus.

There must be some intermediate number $\mu$ between $m$ and $M$ such that

$$\int_a^b f(x)dx = \mu(b-a).$$

If $f(x)$ is continuous and therefore takes all values between $m$ and $M$, it follows that there must exist some point $x_0$ in the interval $(a, b)$ for which

$$f(x_0) = \mu.$$

We have then the following theorem.

THEOREM IX. *If $f(x)$ is continuous in the closed interval $(a, b)$, then there exists at least one point $x_0$ in this interval for which*

$$\int_a^b f(x)dx = f(x_0)\,(b-a).$$

By a similar argument we can prove the following theorem. The details of the proof are left to the reader to supply.

THEOREM X. *If $\phi(x)$ is continuous and $f(x)$ is a bounded, integrable function and non-negative for all values of $x$ in the interval $(a, b)$, then at least one point $x_0$ can be found in this interval for which we have*

$$\int_a^b f(x)\phi(x)dx = \phi(x_0)\int_a^b f(x)dx.$$

The following theorem gives a relation known as the **Second Law of the Mean** for the integral calculus.[1]

---

[1] This form of the Second Law of the Mean is due to Bonnet, see *Liouville's Journal*, Vol. 14 (1849), p. 249.

THEOREM XI. *Let $f(x)$ be bounded and integrable in the interval $(a, b)$ and suppose $\phi(x)$ to be a non-negative, bounded, monotone decreasing function. Then there exists a value $\xi$ such that*

$$\int_a^b f(x)\phi(x)dx = \phi(a)\int_a^\xi f(x)dx,\ a \leqq \xi \leqq b.$$

It follows from the conditions of the theorem that $\phi(x)$ is integrable and hence that the product $f(x)\cdot\phi(x)$ is integrable by Theorem III of Art. 40. The integral $\int_a^b f(x)\phi(x)dx$ is the limit, as the norm $\Delta$ of the intervals $\Delta_k x \equiv (x_k - x_{k-1})$ approaches zero and consequently $n$ becomes infinite, of the sum

$$\sum_{k=1}^{n} f(x_{k-1})\phi(x_{k-1})\,(x_k - x_{k-1}), \tag{1}$$

where $x_0 \equiv a$, $x_n \equiv b$. For convenience put

$$f(x_{k-1})\,(x_k - x_{k-1}) \equiv \lambda_{k-1},$$
$$\phi(x_{k-1}) \equiv \phi_{k-1},\ \lambda_0 + \lambda_1 + \cdot\ \cdot\ \cdot + \lambda_k = A_k.$$

The sum (1) may then be written

$$\begin{aligned}\sum_{k=1}^{n} f(x_{k-1})\phi(x_{k-1})(x_k - x_{k-1}) &= \lambda_0\phi_0 + \lambda_1\phi_1 + \lambda_2\phi_2 + \cdot\ \cdot\ \cdot + \lambda_{n-1}\phi_{n-1}\\ &= (A_0\phi_0 + A_1\phi_1 + A_2\phi_2 + \cdot\ \cdot\ \cdot + A_{n-1}\phi_{n-1})\\ &\quad -(A_0\phi_1 + A_1\phi_2 + \cdot\ \cdot\ \cdot + A_{n-2}\phi_{n-1})\\ &= \sum_{k=1}^{n-1} A_{k-1}(\phi_{k-1} - \phi_k) + A_{n-1}\phi_{n-1}. \qquad (2)\end{aligned}$$

Denote by $\alpha$ the largest and by $\beta$ the smallest of the numbers $A_0, A_1, \cdot\ \cdot\ \cdot, A_{n-1}$. The above summation would be increased should we replace $A_{k-1}$ by $\alpha$ and it would be likewise decreased if we replace $A_{k-1}$ by $\beta$. By making these substitutions, the right-hand member of the above equation becomes either

$$\alpha\left\{\sum_{k=1}^{n-1} (\phi_{k-1} - \phi_k) + \phi_{n-1}\right\} = \alpha\phi_0,$$

or

$$\beta\left\{\sum_{k=1}^{n-1} (\phi_{k-1} - \phi_k) + \phi_{n-1}\right\} = \beta\phi_0.$$

Consequently, we have from (2)

$$\beta\phi_0 \leqq \sum_{k=1}^{n} f(x_{k-1})\phi(x_{k-1})\,(x_k - x_{k-1}) \leqq \alpha\phi_0.$$

Recalling the values of $\alpha$ and of $\beta$, we can now say that, upon passing to the limit as $\Delta$ approaches zero, the integral

$$\int_a^b f(x)\phi(x)dx \tag{3}$$

lies between the least upper and the greatest lower bounds of $\phi(a)\int_a^X f(x)dx$, where $a \leqq X \leqq b$. Since the integral $F(X) \equiv \int_a^X f(x)dx$ is a continuous function[1] of $X$, then $F(X)$ takes all values between its least upper and its greatest lower bounds. Consequently, there is some value of $X$, say $X = \xi$, for which $F(X)$ is exactly equal to the integral (3). We have then

$$\int_a^b f(x)\phi(x)dx = \phi(a)\int_a^\xi f(x)dx, a \leqq \xi \leqq b,$$

as the theorem states.

In the foregoing proof, we have placed upon $\phi(x)$ the restriction that it shall be a bounded, non-negative, monotone decreasing function. If $\phi(x)$ is a bounded, non-negative, monotone increasing function, we can by a similar argument show that

$$\int_a^b f(x)\phi(x)dx = \phi(b)\int_\xi^b f(x)dx, a \leqq \xi \leqq b.$$

Another form of the foregoing theorem has been given by Weierstrass and Du Bois Reymond,[2] namely:

$$\int_a^b f(x)\phi(x)dx = \phi(a)\int_a^\xi f(x)dx + \phi(b)\int_\xi^b f(x)dx, a \leqq \xi \leqq b.$$

It is to be noted that this second Law of the Mean holds even if $f(x)$ is not bounded but merely any integrable function.[3]

Let us consider what changes, if any, need be made in the law for the **change of variables** in order that it may hold for the class of bounded, integrable functions. In this connection we shall prove the following theorem.

THEOREM XII. *Let $f(x)$ be bounded in the interval $(a, b)$ and suppose $x = \phi(t)$ to be a monotone, continuous function of $t$ in the closed interval $(\alpha, \beta)$, where $a = \phi(\alpha)$, $b = \phi(\beta)$. Then, if $D\phi(t)$ denotes any*

---

[1] Theorem I, Art 41.

[2] See *Crelle's Journal*, Vol. 49 (1869), p. 81.

[3] Compare Pringsheim, *Münch. Ber.*, Vol. 30 (1900), p. 218; Hobson, *London Math. Soc.* (2), Vol. 7 (1908), p. 14.

*one of the derivative numbers of* $\phi(t)$ *and if it is bounded and integrable in* $(a, b)$, *we have*

$$\overline{\int_a^b} f(x)dx = \overline{\int_\alpha^\beta} f[\phi(t)]\cdot D\phi(t)dt, \qquad (a)$$

$$\underline{\int_a^b} f(x)dx = \underline{\int_\alpha^\beta} f[\phi(t)]\cdot D\phi(t)dt. \qquad (b)$$

We shall prove $(a)$; the proof for $(b)$ is similar and can be readily supplied. Since $f(x)$ is bounded, the upper integral $\overline{\int_a^b} f(x)dx$ exists. As $\phi(t)$ is continuous in $(\alpha, \beta)$ and $f(x)$ is bounded, $f[\phi(t)]$ is bounded. $D\phi(t)$ is bounded by hypothesis. Consequently, the product $f[\phi(t)]\cdot D\phi(t)$ is bounded and its upper integral also exists. We need to show the equality of the two limits

$$\underset{\Delta x\to 0}{L}\ \bar{S}_x \equiv \underset{\Delta x\to 0}{L}\ \Sigma\, M_x\Delta_k x,\quad \underset{\Delta t\to 0}{L}\ \bar{S}_t \equiv \underset{\Delta t\to 0}{L}\ \Sigma\, M_t\Delta_k t,$$

where $M_x \equiv Max\, f(x)$ in $\Delta_k x$, $M_t \equiv Max\, f[\phi(t)]\cdot D\phi(t)$ in $\Delta_k t$.

As $x = \phi(t)$ is a continuous function, $\Delta_k x$ approaches zero whenever $\Delta_k t$ approaches zero.

As we have seen, the least upper and the greatest lower bounds of the derivative numbers are the same and also the same as the least upper and the greatest lower bounds of the difference quotient. For convenience, put

$$F(t) \equiv \frac{\phi(t) - \phi(t_0)}{t - t_0},\ |\, t - t_0 \,| \leqq \Delta_k t.$$

Denote the least upper and the greatest lower bounds of $F(t)$ in any interval $\Delta_k t$ by $\mu_k$ and $\lambda_k$, respectively. From the integrability of $D\phi(t)$, it follows that by a proper choice of the norm $\Delta t$ we have

$$\Sigma(\mu_k - \lambda_k)\ \Delta_k t < \epsilon,$$

where $\epsilon$ is arbitrarily small.

We have also the relations

$$M_t \geqq Max\, f[\phi(t)]\cdot\lambda_k = M_x\lambda_k,$$

$$\Delta_k x = x_k - x_{k-1} = \frac{\phi(t_k) - \phi(t_{k-1})}{t_k - t_{k-1}}\,\Delta_k t$$

Then we have
$$\begin{aligned} |\,\bar{S}_x - \bar{S}_t\,| &= |\,\Sigma\, M_x\Delta_k x - \Sigma\, M_t\Delta_k t\,| \\ &\leqq |\,\Sigma\, M_x\mu_k\Delta_k t - \Sigma\, M_x\lambda_k\Delta_k t\,| \\ &\leqq |\,\Sigma\, M_x(\mu_k - \lambda_k)\Delta_k t\,| \\ &\leqq Max\,|\,f(x)\,|\,\Sigma(\mu_k - \lambda_k)\,\Delta_k t \\ &\leqq Max\,|\,f(x)\,|\cdot\epsilon \leqq \epsilon_1, \end{aligned}$$

where $\epsilon_1$ is an arbitrarily small positive number, and $Max\,|f(x)|$ is the least upper bound of $|f(x)|$ in $(a, b)$. Consequently, as the norm $\Delta t$, and hence $\Delta x$, approaches zero we have

$$\underset{\Delta t\to 0}{L}\; |\bar{S}_x - \bar{S}_t| = 0$$

and therefore
$$\overline{\int_a^b} f(x)dx = \overline{\int_\alpha^\beta} f[\phi(t)]D\phi(t)dt.$$

We can now establish the following theorem:

THEOREM XIII. *Let $x = \phi(t)$ be a monotone continuous function of $t$ in the closed interval $(\alpha, \beta)$, where $a = \phi(\alpha)$, $b = \phi(\beta)$. Then if one of the derivative numbers $D\,\phi(t)$ is integrable in $(\alpha, \beta)$ and if either of the integrals*

$$\int_a^b f(x)dx, \quad \int_\alpha^\beta f[\phi(t)]D\phi(t)dt \tag{a}$$

*exists, the other also exists and the two are equal.*

From Theorem XII, we have

$$\overline{\int_a^b} f(x)dx = \overline{\int_\alpha^\beta} f[\phi(t)]D\phi(t)dt. \tag{b}$$

$$\underline{\int_a^b} f(x)dx = \underline{\int_\alpha^\beta} f[\phi(t)]D\phi(t)dt \tag{c}$$

By hypothesis, one of the integrals in $(a)$ exists, and hence either the left-hand or the right-hand integrals in $(b)$ and $(c)$ are equal. It follows that the remaining upper and lower integrals in $(b)$ and $(c)$ are equal and the theorem follows.

If the derivative $\phi'(t)$ exists for each value of $t$ in $(\alpha, \beta)$, then of course the derivative number $D\phi(t)$ exists. Moreover, if $\phi'(t)$ is finite, continuous and different from zero at each point, then $\phi(t)$ is monotone. Hence, we have the following corollary.

COR. *Let $f(x)$ be a bounded, integrable function, and suppose $x = \phi(t)$ to be a continuous function of $t$ having a finite, continuous derivative different from zero at all points of the closed interval $(\alpha, \beta)$. We may then write*

$$\int_a^b f(x)dx = \int_\alpha^\beta f[\phi(t)]\,\phi'(t)\,dt.$$

The foregoing theorems involve the condition that $\phi(t)$ be a monotone function. If it is not, the given interval can often be broken up into subintervals where this condition holds and thus the application of the theorems extended.

**41. Indefinite integrals.** Suppose that $f(x)$ is bounded and integrable in the interval $(a, b)$. Then by Theorem I of the last article, the integral $\int_a^x f(x)dx$ exists, where $x$ is any intermediate value between $a$ and $b$. This integral is called the **indefinite integral** and defines a function of $x$. We may write

$$\phi(x) \equiv \int_a^x f(x)dx.$$

The indefinite integral is often written $\int f(x)dx$ without indicating the limits of integration. The indefinite integral has certain properties as expressed in the following theorems.

THEOREM I. *If $f(x)$ is bounded and integrable in the interval $(a, b)$, then*

$$\phi(x) = \int_a^x f(x)dx$$

*is a continuous function of $x$ in this interval.*

By giving to $x$ an increment $\Delta x$ we have

$$\begin{aligned}\phi(x + \Delta x) - \phi(x) &= \int_a^{x+\Delta x} f(x)dx - \int_a^x f(x)dx \\ &= \int_x^{x+\Delta x} f(x)dx. \qquad (1)\end{aligned}$$

We have by the first Law of the Mean

$$\left| \int_x^{x+\Delta x} f(x)dx \right| \leqq M \cdot \Delta x, \qquad (2)$$

where $M$ is the least upper bound of $|f(x)|$ in the interval $(x, x + \Delta x)$. From (1) and (2) we have

$$|\phi(x + \Delta x) - \phi(x)| \leqq M \cdot \Delta x. \qquad (3)$$

We may put
$$\eta < \frac{\epsilon}{M},$$

where $\epsilon$ is an arbitrarily small positive number. For all values of $\Delta x$ smaller than $\eta$, we then have

$$|\phi(x + \Delta x) - \phi(x)| < \epsilon,$$

which shows that $\phi(x)$ is continuous at each point $x$ in the interval $(a, b)$.

A similar argument shows that the upper and the lower integrals

$$\overline{\int_a^x} f(x)dx, \quad \underline{\int_a^x} f(x)dx$$

are also continuous functions of $x$.

The question naturally arises as to the conditions under which $\phi(x)$, as defined by Theorem I, has a derivative and, if it does exist, what its relations are to $f(x)$. In answer to this question we have the following theorem.

THEOREM II. *If $f(x)$ is bounded and integrable in the interval $(a, b)$, then $\phi(x)$, defined by the relation*

$$\phi(x) \equiv \int_a^x f(x)dx,$$

*has a derivative at each point $x_0$ of $(a, b)$ where $f(x)$ is continuous. Moreover, we have*

$$\left.\frac{d\phi}{dx}\right]_{x=x_0} = f(x_0).$$

Since $x_0$ is a point of continuity of $f(x)$, it follows that for every positive value of $\epsilon$ however small there exists a $\delta > 0$ such that we have

$$|f(x) - f(x_0)| < \epsilon, \; |x - x_0| < \delta. \tag{4}$$

From the definition of $\phi(x)$, we have

$$\phi(x_0 + \Delta x) - \phi(x_0) = \int^{x_0+\Delta x} f(x)dx. \tag{5}$$

In the interval $(x_0 - \delta, x_0 + \delta)$ the least upper bound of $f(x)$ is less than $f(x_0) + \epsilon$, while the greatest lower bound is greater than $f(x_0) - \epsilon$. Hence, from Theorem VIII of the previous article, we have for all values of $|\Delta x| < \delta$

$$\Delta x\,[f(x_0) - \epsilon] \leqq \int_{x_0}^{x_0+\Delta x} f(x)dx \leqq \Delta x\,[f(x_0) + \epsilon].$$

By aid of (5) we may then write

$$\Delta x\,[f(x_0) - \epsilon] \leqq [\phi(x_0 + \Delta x) - \phi(x_0)] \leqq \Delta x\,[f(x_0) + \epsilon],$$

whence

$$f(x_0) - \epsilon \leqq \frac{\phi(x_0 + \Delta x) - \phi(x_0)}{\Delta(x)} \leqq f(x_0) + \epsilon.$$

The first and last terms of this inequality approach the same limiting value $f(x_0)$ as $\delta$ approaches zero. But $\Delta x$ approaches zero with $\delta$. Hence, by Theorem IV, Art. 24, the following limit exists, namely we have

$$\underset{\Delta x \to 0}{L} \frac{\phi(x_0 + \Delta x) - \phi(x_0)}{\Delta x} = f(x_0).$$

But this limit is nothing else than the derivative of $\phi(x)$ at the point $x_0$, and therefore we have

$$\left.\frac{d\phi}{dx}\right]_{x=x_0} = f(x_0).$$

In a similar manner it may be shown that if $f(x)$ is continuous at any point $x_0$ of $(a, b)$, then the integrals

$$\overline{\int_a^x} f(x)dx, \quad \underline{\int_a^x} f(x)dx$$

define functions, each of which has a derivative at $x_0$.

If $f(x)$ is continuous throughout the interval $(a, b)$, then we have as a consequence of this theorem the following corollary.

COR. *If $f(x)$ is continuous in the closed interval $(a, b)$, we have*

$$\phi(x) = \int_a^x f(x)dx, \quad a \leqq x \leqq b,$$

*where, for all values of $x$ in $(a, b)$, we have*

$$\frac{d\phi}{dx} = f(x);$$

*that is, every continuous function is the derivative of some other function.*

It follows from the foregoing theorem that the points at which an indefinite integral $\phi(x)$ can fail to have a derivative form at most a set of points of measure zero; because, these points must be points of discontinuity of the given function $f(x)$, which is by hypothesis a bounded, integrable function.

In case $f(x)$ is bounded and integrable in $(a, b)$, but has points of discontinuity, the right-hand derivative and the left-hand derivative of $\phi(x)$ exist and are equal respectively to $f(x + 0)$ and $f(x - 0)$. The proof of this statement follows the same line of argument as that for the proof of the foregoing theorem and is left for the reader to supply.[1]

A function $F(x)$ which has $f(x)$ as its derivative for all values of $x$ in the interval $(a, b)$ of integration is called a **primitive** of $f(x)$. Any other function which differs from $F(x)$ by only an additive constant is also a primitive of $f(x)$, since its derivative is equal to $f(x)$. If the indefinite integral $\phi(x)$ has a derivative such that

$$\frac{d\phi}{dx} = f(x),$$

then $\phi(x)$ is a primitive of $f(x)$ and we have

$$F(x) = \phi(x) + c.$$

---

[1] See *Encyclop. des Sci. Math.* $\mathrm{II}_2$, p. 199.

This is always true if $f(x)$ is a continuous function, in which case we frequently omit the limits of integration and write

$$\int f(x)dx = F(x) = \phi(x) + c.$$

If $f(x)$ has points of discontinuity but is bounded and integrable, the same statement can not always be made. It may happen, for example, that the indefinite integral $\phi(x)$ may exist and yet there may be values of $x$ for which the derivative $\frac{d\phi}{dx}$ does not exist. It may also happen, as we shall see, that

$$\frac{dF}{dx} = f(x),\ a \leqq x \leqq b$$

and yet $f(x)$ be not integrable in the interval $(a, b)$.

In Arts. 29 and 31 we have discussed briefly the properties of the derivative numbers of a function $f(x)$, represented by $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$, respectively. We shall now consider some of the relations of these numbers to the indefinite integral.

THEOREM IV. *Given any function $\phi(x)$ which is known to be continuous in $(a, b)$ and suppose that one of its derivative numbers is bounded and integrable in $(a, b)$, then the other derivative numbers are also bounded and integrable. Moreover, for any value of $x$ in $(a, b)$ we have*

$$\phi(x) - \phi(a) = \int_a^x D^+\phi(x)dx = \int_a^x D_+\phi(x)dx$$
$$= \int_a^x D^-\phi(x)dx = \int_a^x D_-\phi(x)dx.$$

By the insertion of the intermediate points $x_1, x_2, \cdots x_{n-1}$, let the interval $(a, x)$ be divided into subintervals. For convenience put

$$\Delta_k x \equiv x_k - x_{k-1},\ k = 1, 2, 3, \cdots n, \text{ where } a \equiv x_0,\ x \equiv x_n.$$

Denote by $D\phi(x)$ that derivative number which is known to be integrable in $(a, b)$. Then by Theorem VI of Art. 31, it follows that $\phi(x_k) - \phi(x_{k-1})$ lies between the least upper and the greatest lower bounds of $\Delta_k x D\phi(x)$. If we denote the least upper and the greatest lower bounds of $D\phi(x)$ by $M$ and $m$, respectively, then we may write

$$m_k\Delta_k x \leqq \phi(x_k) - \phi(x_{k-1}) \leqq M_k\Delta_k x.$$

By taking the sum of all such products for the interval $(a, x)$, we have

$$\sum_{k=1}^{n} m_k\Delta_k x \leqq \phi(x) - \phi(a) \leqq \sum_{k=1}^{n} M_k\Delta_k x. \tag{7}$$

From the fact that the one derivative number is known to be bounded and integrable, it follows that the limits of the first and the last terms in (7) must lead to the same limiting value as the norm $\Delta$ of the $\Delta_k x$'s approaches zero and hence $n$ increases indefinitely (Theorem IV, Art. 24), namely the integral of the given derivative number. As the inequality holds for all the derivative numbers of $\phi(x)$, we have he general conclusion desired, namely

$$\phi(x) - \phi(a) = \int_a^x D^+\phi(x)dx = \int_a^x D_+\phi(x)dx$$
$$= \int_a^x D^-\phi(x)dx = \int_a^x D_-\phi(x)dx.$$

If we have

$$D^+\phi(x) = D_+\phi(x) = D^-\phi(x) = D_-\phi(x) = f(x),$$

the function $\phi(x)$ is a primitive of $f(x)$. We then have as a special case of the foregoing theorem the following corollary, known as the **fundamental theorem of integral calculus.**

COR. 1. *If in the interval* $(a,b)$, $f(x)$ *is bounded and integrable and if there exists a function* $F(x)$ *such that* $\frac{dF}{dx} = f(x)$ *for all values of* $x$ *in* $(a\ b)$, *then*

$$\int_a^x f(x)dx = F(x) - F(a).$$

If any one of the derivative numbers $D\phi$ exists and is equal to $f(x)$, we have also the following corollary.

COR. 2. *If* $\phi(x)$ *is a continuous function having a derivative number* $D\phi(x)$ *equal to* $f(x)$ *for all values of* $x$ *in the interval* $(a, b)$, *then*

$$\phi(x) - \phi(a) = \int_a^x f(x)dx.$$

If in the foregoing theorem the given derivative number $D\phi(x)$ is bounded but not integrable, it still follows that the upper and the lower integrals of $D\phi(x)$ exist, and we have the following theorem.

THEOREM V. *Given a bounded, integrable function* $f(x)$ *having the indefinite integral*

$$\phi(x) = \int_a^x f(x)dx.$$

*Then, if* $D\phi(x)$ *is any one of the derivative numbers of* $\phi(x)$, *it follows that*

$$\overline{\int_a^x} D\phi(x)dx \geqq \phi(x) - \phi(a) \geqq \underline{\int_a^x} D\phi(x)dx,\ a \leqq x \leqq b.$$

This theorem follows as an immediate consequence of (7); for, that inequality still holds even if $D\phi(x)$ is not integrable. If, however, it is bounded, then the upper integral and the lower integral both exist and we have in the limit the desired result.

As a special case, the foregoing results hold if $\phi(x)$ has a derivative; that is, when the derivative numbers of $\phi(x)$ are equal. We have then the following corollary.

COR. *If* $\phi(x)$, *defined by the indefinite integral*

$$\phi(x) = \int^x f(x)dx,$$

*has a derivative for each value in the interval* $(a, x)$, *then*

$$\overline{\int_a^x} \frac{d\phi}{dx} dx \geqq \phi(x) - \phi(a) \geqq \underline{\int_a^x} \frac{d\phi}{dx} dx.$$

If we have given $\phi(x) = \int_a^x f(x)dx$

and $\frac{d\phi}{dx}$ exists at every point in $(a,b)$, it does not follow that $\phi(x)$ is necessarily a primitive of $f(x)$, but from the foregoing discussion we see that $\frac{d\phi}{dx}$ can differ from $f(x)$ at most at a set of points of measure zero.

The foregoing theorems lead us to certain important conclusions as to the relation of differentiation and integration. For the simple functions considered in elementary calculus, they are inverse operations. If we include all functions to which these two processes apply, we shall see that they are not always the inverse of each other. It is desirable to see under what conditions they may be so regarded. Suppose we have a function $\phi(x)$ defined by the relation,

$$\phi(x) = \int_a^x f(x)dx,\ a \leqq x \leqq b. \tag{8}$$

The proposed problem involves the conditions that the given integral shall exist and that we shall have

$$f(x) = \frac{d\phi}{dx}. \tag{9}$$

Conversely, it involves the condition that a given function $F(x)$ shall have at each point of $(a, b)$ a unique derivative such that this derived function shall have the given function $F(x)$ as an indefinite integral.

Let us assume $f(x)$ to be a bounded, integrable function. From Theorem II it follows that at each point of continuity of $f(x)$ the relation (9) exists. If $f(x)$ is continuous throughout the interval $(a, b)$, it follows that $f(x)$ is the derivative of the indefinite integral for all values of $x$ in $(a, b)$. In other words, differentiation is in this case the inverse of integration. However, if we assume merely the existence of the indefinite integral $\int_a^x f(x)dx$, rather than the continuity of $f(x)$, it defines a continuous function $\phi(x)$ whose derivative is the given function $f(x)$ for all values of $x$ with the exception of at most a set of points $E$ of measure zero. The points of $E$ where the inverse operation, namely differentiation, fails are points of discontinuity of $f(x)$. As the integral of $f(x)$ exists, the points of discontinuity of $f(x)$ where the saltus exceeds an arbitrary number $\eta$ can not be in any subinterval of $(a, b)$ everywhere dense, and consequently by Theorem III, Art. 28, $f(x)$ can be at most point-wise discontinuous. It follows then from Theorem IV, Art. 28, that the points at which differentiation may be regarded as the inverse of integration must form a set having the cardinal number of the continuum. The points $E$ where the operation of differentiation fails must be a set of at most measure zero. As we have seen, a set of measure zero may be everywhere dense, as for example the rational numbers. Such a set may even have the cardinal number of the continuum. In Art. 20 we formed a set having this property. At all of the points of $E$ the function $\phi(x)$ is continuous by Theorem I, but as we have seen it fails to have a derivative or, if it does have one, that derivative may be different from $f(x)$. It is important to note that if a given function $f(x)$ is integrable in an interval $(a, b)$, then the indefinite integral $\int_a^x f(x)dx$, $a \leqq x \leqq b$, can not fail to define a function $\phi(x)$ which is continuous in $(a, b)$ and has points in every subinterval, however small, where its derivative is equal to $f(x)$.

Let us now consider the inverse problem. Suppose the function $F(x)$ has a derivative $f(x)$ at each point of a given interval $(a, b)$. Does the integral $\int^x f(x)dx$ exist and give $F(x)$ except for an additive constant? From the existence of the derivative at each point of $(a, b)$, it follows that $F'(x) \equiv f(x)$ is at most point-wise discontinuous. In order that $F'(x)$ should satisfy the conditions of integrability it is necessary that it be discontinuous at most at a set of points of measure zero. Such a function need not exist.

Volterra has given the following illustration[1] of a function having a bounded derivative at every point of a given interval $(a, b)$ where this derived function does not satisfy the conditions of integrability in the Riemann sense.

Let $E$ be a perfect set of measure greater than zero in the interval $(a, b)$ but in no portion of it everywhere dense. There must then be in every subinterval a portion which is free of points of $E$. Let $(\alpha, \beta)$ be any one of those subintervals complementary to the set $E$. We shall make use of the function

$$\begin{aligned}\varphi(x, \alpha) &= (x-\alpha)^2 \sin \frac{1}{x-\alpha}, \text{ for } x \neq \alpha,\\ &= 0, \text{ for } x = \alpha.\end{aligned}$$

Differentiating, we have

$$\begin{aligned}\varphi'(x, \alpha) &= 2(x-\alpha) \sin \frac{1}{x-\alpha} - \cos \frac{1}{x-\alpha}, \text{ for } x \neq \alpha,\\ &= 0, \text{ for } x = \alpha.\end{aligned}$$

The derived function $\varphi'(x, \alpha)$ has an infinite number of zeros in the interval $\left(\alpha, \frac{\alpha+\beta}{2}\right)$, and these zero points are dense at $x = \alpha$. Let $\alpha + \gamma$ be the largest value of $x$ for which $\varphi'(x, \alpha)$ is zero in $\left(\alpha, \frac{\alpha+\beta}{2}\right)$. We shall now define $f(x)$ as follows:

$$\begin{aligned}f(x) &= \varphi(x, \alpha), \text{ for } \alpha < x \leqq \alpha+\gamma\\ &= \varphi(\alpha+\gamma, \alpha), \text{ for } \alpha+\gamma \leqq x \leqq \beta-\gamma\\ &= \varphi(x, \beta), \text{ for } \beta-\gamma \leqq x < \beta\\ &= 0, \text{ for all points of } E.\end{aligned} \tag{10}$$

It follows that $f(x)$ is fully defined and continuous in the interval $(a, b)$. Since $E$ is a perfect set it must include its limiting points, and hence $\alpha$, $\beta$ belong to $E$.

At each point $x_0$ of $E$ the derivative $f'(x)$ is zero, as we shall see. Form the difference quotient

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

As $x_0$ is a point of $E$, $f(x_0)$ is zero. If $x_0 + \Delta x$ is likewise a point of $E$, then $f(x_0 + \Delta x)$ is also zero. If, on the other hand, $x_0 + \Delta x$ is a point of the complementary set $C(E)$, then $f(x_0 + \Delta x)$ takes the values given in (10). In any case, the limit of the difference quotient as $\Delta x$ approaches zero is zero. Consequently, for each point of $E$ the derivative is zero. By Theorem IV, Art. 17, it follows that every point of $E$ is an $\alpha$-point or a $\beta$-point or the limit of such points. Since this is the case, each point of $E$ is a point at which $f'(x)$ is zero but in the neighborhood of which this derived function oscillates between $+1$ and $-1$. Each point of $E$ is therefore a point of discontinuity of $f'(x)$ having the saltus 2. As $E$ has by hypothesis a measure greater than zero, $f'(x)$ can not be integrable in the Riemann sense.

---

[1] See *Giornali de Mat.* (1), Vol. 19 (1881), pp. 333–72; also *Encyclop. Sci. Math.* II$_2$, p. 195. For another illustration, see Hahn, *Monatshefte für Math. und Physik.* Vol. 16 (1905), pp. 161–166.

From this illustration it follows that integration is not necessarily the inverse of differentiation if we place no farther restriction upon the derivative than its existence.

If we assume that the derivative function

$$F'(x) \equiv f(x)$$

is bounded and has only ordinary finite discontinuities in the given interval $(a, b)$, then the indefinite integral $\int_a^x f(x)dx$ exists and is equal to $F(x)$ except for an additive constant; that is, integration is then the inverse of differentiation. For, from Theorem VIII of Art. 39, it follows that the integral $\int_a^x f(x)dx$ exists for all values of $x$ where $a \leqq x \leqq b$. As we have seen,[1] the right-hand derivative and the left-hand derivative must exist in this case at each of these points of discontinuity and must be equal to $f(x+0)$ and $f(x-0)$, respectively; that is, they must be equal to $F'(x+0)$, $F'(x-0)$. But as $F'(x)$ exists at each point of $(a, b)$, these two values are the same. Consequently, under the conditions stated, it follows that integration is the inverse of differentiation.

In case a function $F(x)$ has a continuous derivative in $(a, b)$, it follows from the foregoing discussion that this derivative is integrable and that integration gives the original function as the inverse operation requires.

The only case where a given derived function $F'(x)$ can fail to have an indefinite integral which differs from $F(x)$ at most by an additive constant, that is the only case where integration fails to be an inverse operation to differentiation, is where $F'(x)$ has points of discontinuity in every neighborhood of which $F'(x)$ has an infinite number of oscillations. The example given above is an illustration. As a matter of fact, a continuous function $F(x)$ which oscillates everywhere can not have a derivative function $Df(x)$ which is integrable. For, in any subinterval, the least upper bound of the derivative numbers will be positive while the greatest lower bound is negative. Consequently, if we find the limit which defines the upper integral of the derivative function, it is positive while the limit defining the lower integral is negative. As the upper and the lower integrals can not be equal in any subinterval $(a, x)$, it follows that the indefinite integral $\int_a^x f(x)dx$ can not exist.

---

[1] See remarks following Theorem II of this article.

**42. Definition of infinite integrals.** Thus far we have discussed integrals of bounded functions over a finite interval. We shall now consider what extensions may be made to cases where either the integrand becomes infinite at certain points or one or both of the limits of integration become infinite. In contradistinction to those integrals already discussed, we shall designate this new class of integrals as infinite integrals. As we shall see, they appear as limiting cases of finite integrals.

Let us suppose that $f(x)$ has an infinite discontinuity at a point $x_0$ in the interval of integration $(a,b)$, and that at all other points of $(a,b)$ $f(x)$ is bounded. We can exclude the point $x_0$ by inclosing it in a small interval $(x_0-\epsilon_1, x_0+\epsilon_2)$ where $\epsilon_1, \epsilon_2$ are arbitrarily chosen. In the remaining portion of $(a,b)$, $f(x)$ is bounded, and the two integrals

$$\int_a^{x_0-\epsilon_1} f(x)dx, \int_{x_0+\epsilon_2}^b f(x)dx$$

exist under the conditions set forth in the foregoing articles. The above integrals are functions of $\epsilon_1$, $\epsilon_2$, respectively, and it may happen that the two limits

$$\underset{\epsilon_1 \to 0}{L} \int_a^{x_0-\epsilon_1} f(x)dx, \ \underset{\epsilon_2 \to 0}{L} \int_{x_0+\epsilon_2}^b f(x)dx \tag{1}$$

both exist. In this case we call the sum of these limiting values the **infinite integral** [1] of $f(x)$ taken over the interval $(a, b)$ and denote it by the symbol

$$\int_a^b f(x)dx.$$

It is essential to notice that $\epsilon_1$, $\epsilon_2$ are independent of each other. If this were not so, it would be possible for the sum of the two integrals in (1) to have a limit although each limit might not exist. For example, consider the integral of $\frac{1}{x^3}$ over the interval $(-1, 1)$. This function becomes infinite at $x = 0$. If we put $\epsilon_1 = \epsilon_2 = \epsilon$, we have

$$\underset{\epsilon \to 0}{L} \left[ \int_{-1}^{-\epsilon} \frac{dx}{x^3} + \int_{\epsilon}^{1} \frac{dx}{x^3} \right] = \underset{\epsilon \to 0}{L} \left[ -\frac{1}{2} \left\{ \frac{1}{\epsilon^2} - 1 + 1 - \frac{1}{\epsilon^2} \right\} \right] = 0.$$

---

[1] Infinite integrals are sometimes called *improper integrals.* The term *infinite integral* was first introduced by Hardy (see *Proc. London Math. Soc.*, Vol. 34 (1901), p. 16). This term is very suggestive in that it points out the distinction between these integrals and finite integrals (definite integrals) in much the same manner as we distinguish between infinite series and finite series. The use of the term convergence of infinite integrals is also suggestive of the same analogy.

However, neither of the limits

$$\underset{\epsilon_1 \to 0}{L} \int_{-1}^{-\epsilon_1} \frac{dx}{x^3}, \quad \underset{\epsilon_2 \to 0}{L} \int_{\epsilon_2}^{1} \frac{dx}{x^3}$$

exists, and consequently the integral $\int_{-1}^{1} \frac{dx}{x^3}$ can not be said to exist.

If the limit $$\underset{\epsilon \to 0}{L} \left[ \int_a^{x_0 - \epsilon} f(x)dx + \int_{x_0+\epsilon}^{b} f(x)dx \right]$$

exists, where we have put $\epsilon_1 = \epsilon_2 = \epsilon$, the result was called by Cauchy the **principal value** of the integral. It will be seen from the foregoing illustration that the principal value may exist without the infinite integral $\int_a^b f(x)dx$ existing.[1]

The definition of an infinite integral can be extended at once to the case where $f(x)$ has a finite number of points of infinite discontinuity. Let $c_1$, $c_2$ be any two such points which are consecutive. In the interval $(c_1 + \epsilon_1, c_2 - \epsilon_2)$, $f(x)$ is bounded and let us suppose it to be integrable; that is, the integral

$$\int_{c_1+\epsilon_1}^{c_2-\epsilon_2} f(x)dx$$

exists. We then say that the infinite integral $\int_{c_1}^{c_2} f(x)dx$ exists if the limit

$$\underset{\substack{\epsilon_1 \to 0 \\ \epsilon_2 \to 0}}{L} \int_{c_1+\epsilon_1}^{c_2-\epsilon_2} f(x)dx$$

exists. If in each of the divisions of $(a, b)$ made by the points $c_1, c_2, c_3, \cdot \cdot \cdot, c_n$ of infinite discontinuity the corresponding limit exists, then we say that the infinite integral $\int_a^b f(x)dx$ exists.

It will also be seen that this definition can be extended to the case where the set $E$ of points of infinite discontinuity is such that the derived set $E'$ consists of a finite number of points.[2] For, let $\alpha$, $\beta$ be any two consecutive points of $E'$. Then in the interval $(\alpha + \epsilon_1, \beta - \epsilon_2)$ there can be only a finite number of points of $E$, and, as we have seen, the definition can be extended to such cases. If the infinite integral

$$\int_{\alpha+\epsilon_1}^{\beta-\epsilon_2} f(x)dx$$

---

[1] For a discussion of the properties of the principal values of an infinite integral, see Hardy, *Proc. London Math. Soc.*, Vol. 34 (1902), pp. 16–40, also pp. 55–91, and Vol. 35 (1903), pp. 81–107.

[2] Due to Dirichtlet. See Lipschitz, *Journ. für die reine und angewandte Math.*, Vol. 63 (1864), pp. 296–308.

exists also for the interval between each consecutive pair of points $\alpha$, $\beta$ of $E'$ and has a definite limit as $\epsilon_1$, $\epsilon_2$ approach zero, then the infinite integral $\int_a^b f(x)dx$ is said to exist.

As an illustration of such a function, we have

$$f(x) = \csc\frac{1}{x}, \quad 0 < x < 1,$$

which has an infinite discontinuity at a set of points $E$ dense at the one point $x = 0$; for, writing

$$\csc\frac{1}{x} = \frac{1}{\sin\frac{1}{x}},$$

we see that the points $E$ are those where $\sin\frac{1}{x}$ has zero points. The derived set $E'$ consists of a single point, namely $x = 0$.

We may well question how far the set $E$ of infinite discontinuities may be extended and still preserve the essential characteristic properties of finite integrals. For example, it is desirable that the indefinite integral should be uniquely determined except as to an additive constant in every subinterval where the integrand is bounded and integrable; moreover, it is also desirable that the indefinite integral be continuous and satisfy the fundamental law of the integral calculus.

We may therefore state the definition of an infinite integral of non-bounded functions as follows. Given a function $f(x)$ having a set $E$ of infinite discontinuities in the interval $(a, b)$, we say that $f(x)$ is integrable in $(a, b)$, if there exists in $(a, b)$ a continuous function $\phi(x)$ which is uniquely determined, except as to an additive constant, satisfying the relation

$$\int_{x_0}^{x_1} f(x)dx = \phi(x_1) - \phi(x_0) \tag{1}$$

for every subinterval $(x_0, x_1)$ of $(a,b)$ in which $f(x)$ is bounded and integrable.[1] If these conditions are satisfied, $\phi(x)$ is said to be the indefinite integral of $f(x)$, and we write

$$\int_a^x f(x)dx = \phi(x) - \phi(a), a \leqq x \leqq b.$$

---

[1] See Hölder, *Math. Ann.*, Vol. 24 (1884), p. 192; also Lebesgue, *Leçons sur l'intégration*, p. 13. Other definitions of integrals of non-bounded functions are to be found in the literature. See Hamach, *Math. Ann.*, Vol. 21 (1883), p. 34, Vol. 24, p. 220; de Vallée Poussin, *Jour. de Math.* (4), Vol. 8 (1892), p. 427; E. H. Moore, *Trans. Amer. Math. Soc.*, Vol. 2 (1901), p. 459; Pierpont, *Real Variables*, Vol. 2, p. 30.

In order to state the conditions under which $f(x)$ shall satisfy this generalized definition, it is convenient to introduce what has been called a **reducible set of points,** by which is understood such a set $E$ that some one of the derived sets

$$E', E'', \cdot\ \cdot\ \cdot, E^{(n)}, \cdot\ \cdot\ \cdot, E^{(\omega)}_{a}$$

contains only a finite number of points. Such a set is enumerable by Theorem VIII of Art. 15. Without some qualifications as to the character of the set, the converse is not true. For example, the set of rational numbers is enumerable, but it is not reducible. We may now state the following theorem.

THEOREM I. *In order that $f(x)$ be integrable in the interval $(a, b)$, it is necessary and sufficient that the points of infinite discontinuity form a reducible set and that there exists a continuous function $\phi(x)$ which, in every subinterval $(x_0, x_1)$ where $f(x)$ is bounded and integrable, satisfies the relation* (1).

In order to show that the foregoing conditions are sufficient for the integrability of $f(x)$, it must be shown that throughout the interval $(a, b)$ the indefinite integral $\phi(x)$ may be uniquely determined, except for an additive constant, if the points of infinite discontinuity form a reducible set $E$. We have already seen that $\phi(x)$ may be uniquely determined, except as to an additive constant, if $E$ consists of a finite number of points and likewise if $E'$ consists of a finite number of points.

If the second derived set $E''$ consists of a finite number of points, the same conclusion can be reached for those subintervals free from points of $E''$. Let $c_1, c_2$ be any consecutive points of $E''$. Then in the interval $(c_1 + \epsilon_1, c_2 - \epsilon_2)$, $f(x)$ is integrable and $\phi(x)$ is uniquely determined, except for an additive constant, since, as we have seen, it is true for intervals where $E'$ consists of a finite number of points. Moreover, the limit

$$\underset{\substack{\epsilon_1 \to 0 \\ \epsilon_2 \to 0}}{L} \int_{c_1 + \epsilon_1}^{c_2 - \epsilon_2} f(x)dx$$

must exist because $\phi(x)$ defined by this integral is a continuous function in the closed subinterval $(c_1, c_2)$. It follows in this case that $\phi(x)$ is uniquely determined, except as to an additive constant, throughout the interval $(a, b)$ and the integral $\int_a^b f(x)dx$ satisfies the definition of an infinite integral. In a similar manner, it can be shown that $\phi(x)$ is uniquely determined, except for an additive constant, if any finite derived set $E^{(n)}$ consists of a finite number of points.

Finally, let us assume that $E^{(\omega)}$ consists of a finite number of points, of which $c_1$, $c_2$ are any two consecutive ones. It is then possible to find a value of $n$, such that in the interval $(c_1+\epsilon_1, c_2-\epsilon_2)$ the derived set $E^{(n)}$ consists of a finite number of points. In this interval the integral of $f(x)$ exists and defines uniquely, except as to an additive constant, an indefinite integral $\phi(x)$. From the fact that this conclusion holds for every closed interval free from points of $E^{(\omega)}$ and from the assumed continuity of $\phi(x)$ in these intervals, it follows that $f(x)$ satisfies the definition of an infinite integral in the given interval $(a, b)$.

These conditions are also necessary; for, if $f(x)$ is integrable in accordance with the definition given, then $\phi(x)$ is uniquely determined, except for an additive constant, in any subinterval of $(a, b)$ which is free of points of $E$ and is continuous throughout the interval of integration. Moreover, we have seen that the set $E$ can be reducible for an integrable function, and the following illustration will show that, in case it is non-reducible, the indefinite integral $\phi(x)$ is not necessarily uniquely determined.

The indefinite integral $\phi(x)$ can be determined even though the set $E$ of infinite discontinuities is not reducible; in fact, even if it is a perfect set. However, under these circumstances $\phi(x)$ is not necessarily uniquely determined except as to an additive constant. To show this, let $E$ be the perfect set discussed in Art. 17 formed by dropping out the middle third of successive free intervals. Let $f(x)$ be equal to $x$ in all points of the interval (0, 1) except at points of $E$, where we assume that $f(x)$ has an infinite discontinuity. Since $E$ is a closed set of measure zero, it is also of content zero and hence can be inclosed in a finite number $k$ of intervals, each of arbitrarily small length, say equal to $\epsilon$. We then have

$$\phi(x) = \int_0^x f(x)dx = \underset{\epsilon \to 0}{L} \frac{x^2}{2}(1 - k\,\epsilon) = \frac{x^2}{2}, \; 0 \leqq x \leqq 1.$$

We shall now show that $\phi(x)$ is not uniquely determined except for an additive constant; that is, we shall show that it is possible to define for the interval (0, 1) a function $F(x)$ such that $\phi(x) + F(x)$ is continuous in (0, 1) and likewise satisfies (1) in any subinterval of (0, 1) not containing points of $E$.[1] Let

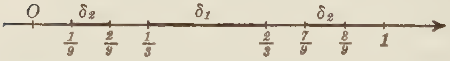


Fig. 33.

---

[1] See Lebesgue, *Leçons sur l'intégration*, p. 13.

$$\begin{aligned} F(0) &= 0;\ F(1) = 1;\\ F(x) &= \tfrac{1}{2},\ \text{for } x \text{ in the open interval } \delta_1;\\ &= \tfrac{1}{4},\ \text{for } x \text{ in the open interval } \delta_2,\ \text{if } 0 < \delta_2 < \delta_1;\\ &= \tfrac{3}{4},\ \text{for } x \text{ in the open interval } \delta_2,\ \text{if } \delta_1 < \delta_2 < 1;\\ &\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\quad \cdot\\ &= \tfrac{1}{2}\,[F(\delta_i) + F(\delta_j)], \end{aligned}$$

where $\delta_i$, $\delta_j$ are the two intervals of the sequence

$$\delta_1,\ \delta_2,\ \delta_3,\ \cdot\ \cdot\ \cdot\ ,\ \delta_{k-1}$$

which lie on opposite sides of $\delta_k$. All points of $E$ are end-points of the $\delta$-intervals or are the limiting points of these end-points. As $x$ approaches any one of the points of $E$, the values of $F(x)$ approach a definite limit, which we assign as the value of $F(x)$ at those points. Thus $F(x)$ is fully defined in the interval $(0, 1)$. It is continuous throughout this interval but is not a constant for all values of $x$ in the interval. It is, however, a constant in any subinterval $\delta$ not containing points of $E$. The function $\psi(x)$, where

$$\psi(x) \equiv F(x) + \phi(x),$$

is then continuous in $(0, 1)$ and satisfies the relation (1). It follows that if the set $E$ of infinite discontinuities is non-reducible, the indefinite integral $\phi(x)$ is not necessarily uniquely determined, except as to an additive constant, in those subintervals not containing points of $E$.

Since every reducible set is enumerable and hence of measure zero, it follows from the foregoing theorem that the points of infinite discontinuity of a non-bounded, integrable function can at most form a set of measure zero.

If **one of the limits of integration is infinite,** then the infinite integral $\int_a^\infty f(x)dx$ is said to exist, provided the limit

$$\underset{x\to\infty}{L} \int_a^x f(x)dx$$

exists, and the integral is defined as that limit. A similar statement can be made when the lower limit of integration is infinite. If both limits of integration are infinite, then, in order that the infinite integral $\int_{-\infty}^\infty f(x)dx$ may be said to exist, it is necessary that the two limits

$$\underset{x\to\infty}{L} \int_{-x}^a f(x)dx,\ \underset{x\to\infty}{L} \int_a^x f(x)dx$$

shall exist.

**Ex. 1.** Evaluate the integral $\int_1^\infty \frac{dx}{1+x^2}$.

We have

$$\int_1^\infty \frac{dx}{1+x^2} = \underset{x\to\infty}{L} \int_1^x \frac{dx}{1+x^2} = \underset{x\to\infty}{L} \left(\text{arc tan } x - \frac{\pi}{4}\right) = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}.$$

**Ex. 2.** Test the existence of the integral $\int_0^\infty \sin x\, dx$.

We have $\int_0^\infty \sin x\, dx = \underset{x\to\infty}{L} \int_0^x \sin x\, dx = \underset{x\to\infty}{L} (1 - \cos x)$.

This limit does not exist. Consequently, the given integral can not be said to exist.

**Ex. 3.** Test the existence of the integral $\int_0^1 \frac{dx}{\sqrt{1-x^2}}$.

The integrand becomes infinite at $x = 1$. However, the integral exists; for, we have

$$\underset{\epsilon\to 0}{L} \int_0^{1-\epsilon} \frac{dx}{\sqrt{1-x^2}} = \underset{\epsilon\to 0}{L} [\text{arc sin } (1-\epsilon) - \text{arc sin } 0] = \underset{\epsilon\to 0}{L} \text{ arc sin } (1-\epsilon) = \frac{\pi}{2}.$$

If the limit defining either of the two types of infinite integrals exists, then the integral is said to **converge.** Otherwise, the integral is **divergent.** If the infinite integral still converges when the absolute value of the integrand is taken; that is, if the infinite integral $\int_a^b |f(x)|\, dx$ converges, we say that the integral **converges absolutely.**

By aid of Theorem VII, Art. 24, we may state a necessary and sufficient condition for the convergence of an infinite integral as follows, since the convergence depends upon the existence of a limit which is equivalent to the inequality given in that theorem.

THEOREM II. *Given a function $f(x)$ defined for the interval $(a, b)$ and having an infinite discontinuity at the point $x_0$, where $a \leqq x_0 \leqq b$. A necessary and sufficient condition that the integral $\int_a^b f(x)dx$ converges is that we have*

$$\underset{\substack{x_1\to x_0\\ x_2\to x_0}}{L} \int_{x_1}^{x_2} f(x)dx = 0,$$

*where $x_2$ is any value lying between $x_1$ and $x_0$.*

A similar condition holds for the convergence of an integral having an infinite limit of integration. It may be stated as follows:

THEOREM III. *A necessary and sufficient condition that the integral $\int_a^\infty f(x)dx$ converges is that the limit $\underset{\substack{x_1\to\infty\\ x_2\to\infty}}{L} \int_{x_1}^{x_2} f(x)dx$ exists and is equal to zero.*

It should be noticed that this theorem does not involve the condition that the function $f(x)$ shall approach zero as $x$ increases without limit,[1] as might be inferred from analogy with the convergence of infinite series, where it is necessary that the limit $\underset{n\to\infty}{L}\, u_n(x)$ shall be zero in order that the series $\Sigma\, u_n(x)$ shall converge. The infinite integral $\int_0^\infty \sin x^2 dx$ is an illustration. The integrand $\sin x^2$ does not approach zero as $x$ increases without limit. However, the integral converges; for, putting $x^2 = t$, we have

$$\int_0^\infty \sin x^2 dx = \int_0^\infty \frac{\sin t}{\sqrt{t}}\, dt,$$

and the latter integral is known to converge.[2]

**43. Convergence of integrals with infinite limits of integration.** Having defined what we shall understand by an infinite integral, we shall now consider some of the properties of those integrals which arise when one or both of the limits of integration become infinite.

In many discussions one is interested primarily in the convergence of such an integral rather than in its actual value. It is important, therefore, that we become familiar with some of the standard tests for convergence. Some of these tests are given in the following theorems.

THEOREM I. *Let $F(x)$ be positive for all values of $x$ in the interval $(a, \infty)$, and suppose that the integral $\int_a^\infty F(x)dx$ converges. If for all values of $x > a$ we have the relation $|f(x)| \leqq F(x)$, then the integral $\int_a f(x)dx$ converges.*

From the given conditions, we have

$$\left| \int_{x_1}^{x_2} f(x)dx \right| \leqq \int_{x_1}^{x_2} F(x)dx, \quad a \leqq x_1 < x_2. \qquad (1)$$

From the convergence of the integral $\int_a^\infty F(x)dx$, we have

$$\underset{\substack{x_1\to\infty \\ x_2\to\infty}}{L} \int_{x_1}^{x_2} F(x)dx = 0.$$

Therefore from (1) it follows that

$$\underset{\substack{x_1\to\infty \\ x_2\to\infty}}{L} \int_{x_1}^{x_2} f(x)dx = 0,$$

and consequently the integral $\int_a^\infty f(x)dx$ converges.

---

[1] This fact was first observed by Dirichlet. See *Journal für die reine und angewandte Mathematik,* Vol. 17 (1837), p. 60.

[2] Cf. Ex. 3, Art. 43.

**Ex. 1.** Test the convergence of the integral $\int_0^\infty \frac{\sin x}{1+x^2}dx$.

By putting $F(x) = \frac{1}{1+x^2}$, we may apply the foregoing theorem. We have, since $|\sin x|$ never exceeds the value one,

$$\left| \int_0^\infty \frac{\sin x}{1+x^2}dx \right| \leqq \int_0^\infty \frac{dx}{1+x^2} = \frac{\pi}{2} \text{ (Ex. 1, Art. 42).}$$

From this conclusion it follows that the integral $\int_0^\infty \frac{\sin x}{1+x^2}dx$ converges.

This test for convergence of an integral may be called the **comparison test** and suggests an analogy to the corresponding test for the convergence of an infinite series of positive terms. The corresponding test for the divergence of an integral may be stated as follows.

THEOREM II. *Let $F(x)$ be positive for all values of $x$ in the interval $(a, \infty)$ and suppose the integral $\int_a^\infty F(x)dx$ to be divergent. Then, if $f(x)$ is always positive or always negative and $|f(x)| \geqq F(x)$ for all values of $x$ greater than some number $x_0 > a$, it follows that the integral $\int_a^\infty f(x)\,dx$ is divergent.*

The proof of this theorem is left to the reader to supply.

**Ex. 2.** Test the convergence or divergence of the integral $\int_{x_0}^\infty \frac{e^x}{x}dx$, $x_0 > 0$.

It is possible to find a value $x_1 > x_0$ such that $\frac{e^x}{x} > e^{\frac{x}{2}}$, for $x > x_1$. But for $x > 0$, the function $e^{\frac{x}{2}}$ is positive, and we have

$$\int_{x_0}^\infty e^{\frac{x}{2}}\,dx = \underset{x\to\infty}{L} \int_{x_0}^x e^{\frac{x}{2}}\,dx = \underset{x\to\infty}{L}\ 2(e^{\frac{x}{2}} - e^{\frac{x_0}{2}}).$$

As this limit does not exist, the integral $\int_{x_0}^\infty e^{\frac{x}{2}}\,dx$ does not converge. Since the given function $\frac{e^x}{x}$ exceeds $e^{\frac{x}{2}}$ for $x > x_1$, it follows from the above theorem that the integral

$$\int_{x_0}^\infty \frac{e^x}{x}dx$$

does not converge.

THEOREM III. *If $f(x)$ is bounded and monotonic and if $\underset{x\to\infty}{L} f(x) = 0$, then the integral $\int_a^\infty f(x)\phi(x)dx$ converges, provided $\left|\int^x \phi(x)dx\right|$ is bounded as $x$ takes all finite values.*

Denote by $M$ the least upper bound of $\left|\int^x \phi(x)dx\right|$ as $x$ takes all finite values. Since $f(x)$ approaches zero monotonically as $x$ increases indefinitely, we can choose a value $x_0$ such that we have for all values of $x \geqq x_0$

$$|f(x)| < \frac{\epsilon}{2M}.$$

By the second Law of the Mean for the integral calculus, we have, for $x_2 > x_1 \geqq x_0$,

$$\left| \int_{x_1}^{x_2} f(x) \cdot \phi(x) dx \right| = \left| f(x_1) \int_{x_1}^{\xi} \phi(x) dx \right|, \; x_2 \geqq \xi \geqq x_1,$$

$$= \left| f(x_1) \left\{ \int_a^{\xi} \phi(x) dx - \int_a^{x_1} \phi(x) dx \right\} \right| \leqq \frac{\epsilon}{2M} \cdot 2M = \epsilon.$$

Hence, we obtain $\displaystyle \mathop{L}\limits_{\substack{x_1 \to \infty \\ x_2 \to \infty}} \int_{x_1}^{x_2} f(x) \phi(x) dx = 0,$

and the given integral converges as the theorem requires.

**Ex. 3.** Test the convergence of the integral $\displaystyle\int_{x_0}^{\infty} \frac{1}{\sqrt{x}} \sin x \, dx$, $x_0 > 0$.

Let $f(x) = \dfrac{1}{\sqrt{x}}\cdot$ Then

$$\mathop{L}\limits_{x \to \infty} f(x) = \mathop{L}\limits_{x \to \infty} \frac{1}{\sqrt{x}} = 0.$$

We have also

$$\left| \int_{x_0}^{x} \varphi(x) dx \right| = \left| \int_{x_0}^{x} \sin x \, dx \right| = \left| \cos x - \cos x_0 \right| \leqq 2.$$

Therefore by the above theorem the given integral $\displaystyle\int_{x_0}^{\infty} \frac{\sin x}{\sqrt{x}} dx$ converges.

This example suggests a somewhat similar test for convergence, which may be stated as follows.

Theorem IV. *If a number $\mu > 1$ exists such that $x^{\mu} \,|\, f(x) \,|$ is less than some finite number $M$ for all values of $x > x_0$, then the integral $\displaystyle\int_a^x f(x) dx$ converges as $x$ becomes infinite.*

Put $$x^{\mu} \,|\, f(x) \,| = \psi(x).$$

We then have

$$\left| \int_{x_0}^{x} f(x) dx \right| \leqq \int_{x_0}^{x} |\, f(x) \,| \, dx = \int_{x_0} \frac{\psi(x)}{x^{\mu}} dx$$

$$\leqq M \int_{x_0} \frac{dx}{x^{\mu}} = \frac{M}{\mu - 1} \left\{ \frac{1}{x_0^{\mu-1}} - \frac{1}{x^{\mu-1}} \right\}.$$

But we have $\displaystyle \mathop{L}\limits_{\substack{x \to \infty \\ x_0 \to \infty}} \frac{M}{\mu - 1} \left\{ \frac{1}{x_0^{\mu-1}} - \frac{1}{x^{\mu-1}} \right\} = 0.$

Consequently, we may write

$$\mathop{L}\limits_{\substack{x \to \infty \\ x_0 \to \infty}} \int_{x_0} f(x) dx = 0,$$

and the integral $\displaystyle\int_a^{\infty} f(x) dx$ converges.

**Ex. 4.** An important function known as the gamma function is defined by the relation $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$. Test its convergence.

If $0 < \alpha < 1$, then $\alpha - 1$ is negative, say $\alpha - 1 = -\lambda$. We can select $\mu$ so that $1 < \mu \geqq \lambda$ and thus have

$$x^\mu \,|\, f(x) \,| = x^\mu \cdot |\, x^{\alpha-1} \cdot e^{-x} \,| = \frac{x^\nu}{e^x},\ \nu \geqq 0,$$

which is bounded for all finite values of $x$. Consequently, $\int_0^x x^{\alpha-1} e^{-x} dx$ converges as $x$ becomes infinite in accordance with the foregoing theorem.

If $\alpha = 1$ then the integrand reduces to $\frac{1}{e^x}$ and the given integral converges directly from the definition of convergence.

If $\alpha > 1$, we have, upon putting $\alpha - 1 = \nu$, where $\nu > 0$,

$$\int_0^x x^{\alpha-1}\, e^{-x}\, dx = \int_0^x x^\nu \cdot \frac{1}{e^x}\, dx.$$

Since $\underset{x\to\infty}{L} \frac{x^{k+\nu}}{e^x} = 0$, we know that this integral converges by Theorem I; for, we have

$$\int_0^x x^\nu \cdot \frac{1}{e^x}\, dx < \int_0^x \frac{dx}{x^k},\ 0 < k < 1,$$

and this last integral is convergent as $x$ becomes infinite. Consequently, the given integral converges for all positive values of $\alpha$.

THEOREM V. *The integral* $\int_a^\infty f(x)dx$ *will converge if the integral* $\int_a^\infty |\, f(x) \,|\, dx$ *converges.*

From Theorem VI, Art. 40, we have the condition that

$$\left| \int_{x_1}^{x_2} f(x)dx \right| < \int_{x_1}^{x_2} |\, f(x) \,|\, dx,\ a < x_1 < x_2.$$

From Theorem II, Art. 42, and the convergence of the second of these integrals, we know that

$$\underset{\substack{x_1\to\infty \\ x_2\to\infty}}{L} \left| \int_{x_1}^{x_2} f(x)dx \right| = 0.$$

Consequently, we have $\underset{\substack{x_1\to\infty \\ x_2\to\infty}}{L} \int_{x_1}^{x_2} f(x)dx = 0,$

and the given theorem follows.

**Ex. 5.** Show that $\int^\infty \frac{\cos x}{x^2}\, dx$ converges.

By the foregoing theorem this integral converges if the integral $\int^\infty \left| \frac{\cos x}{x^2} \right| dx$ converges. We have

$$\int^\infty \left| \frac{\cos x}{x^2} \right| dx = \underset{x\to\infty}{L} \int^x \frac{|\cos x\,|}{|\,x^2\,|}\, dx \leqq \underset{x\to\infty}{L} \int^x \frac{dx}{x^2} = \underset{x\to\infty}{L} \left\{ 1 - \frac{1}{x} \right\} = 1.$$

As this integral exists, it follows that the given integral converges.

It should be noted that absolute convergence furnishes a sufficient but not a necessary condition for the convergence of an infinite integral. The following example furnishes an illustration of an integral which converges, but not absolutely.

**Ex. 6.** Test the integral $\int_0^\infty f(x)\,dx$ for absolute convergence, where $f(x)$ is defined as follows,

$$\begin{aligned} f(x) &= 1, \text{ for } 0 \leqq x \leqq 1. \\ &= 0, \text{ for } n-1 < x \leqq n - \frac{1}{n} \\ &= (-1)^{n+1}, \text{ for } n - \frac{1}{n} < x \leqq n, \text{ where } n = 2, 3, 4, \cdots . \end{aligned}$$

The graph of the function is shown in Fig. 34. The required integral is the sum of a series of integrals

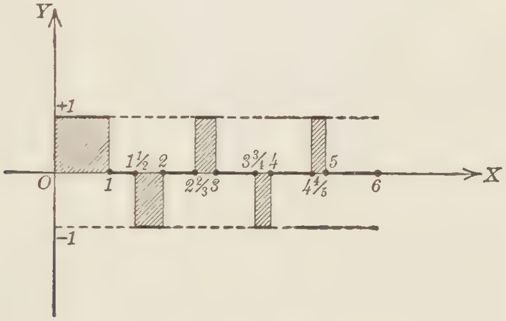


FIG. 34.

$$\begin{aligned} \int_0^\infty f(x)\,dx &= \int_0^1 f(x)\,dx + \int_1^{1\frac{1}{2}} f(x)\,dx + \int_{1\frac{1}{2}}^2 f(x)\,dx + \cdots \\ &= 1 + 0 - \tfrac{1}{2} + 0 + \tfrac{1}{3} + 0 - \tfrac{1}{4} + 0 + \cdots \\ &= 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots . \end{aligned}$$

As we know from elementary mathematics, this series of values converges and hence the integral converges. However, if we had taken the absolute value of $f(x)$, the resulting integrals would have formed a non-convergent series.

**44. Convergence of integrals of non-bounded functions.** We shall now consider the convergence of those infinite integrals $\int_a^b f(x)dx$ where $f(x)$ has one or more points of infinite discontinuity in the interval $(a, b)$. It must not be assumed that the two types of infinite integrals are not related; for, a suitable transformation of the variable will change an infinite integral of the one type to the other as the following examples show.

**Ex. 1.** Express $\int^{\infty} \frac{dx}{x^n}$ as an infinite integral of the second type.

Put $x = \frac{1}{t}$. It follows then that

$$\int_1^{\infty} \frac{dx}{x^n} = -\int_1^0 t^{n-2}\, dt = \int^1 t^{n-2}\, dt.$$

The resulting integral is an infinite integral of the second type if $n < 2$, otherwise it is a finite integral. This example shows also that sometimes it is possible to transform an infinite integral into a finite integral by a suitable transformation.

**Ex. 2.** Transform the infinite integral $\int_0^1 \frac{dx}{\sqrt{1-x^2}}$ into an infinite integral of the first type.

Put $x = \frac{1}{t+1}$. Hence, $dx = -\frac{dt}{(t+1)^2}$ and we have

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} = -\int_{\infty}^0 \frac{dt}{(t+1)^2} \cdot \frac{t+1}{\sqrt{t^2+2t}} = \int_0^{\infty} \frac{dt}{(t+1)\sqrt{t^2+2t}}.$$

It is to be noted that both the given integral and the resulting integral are convergent.

The tests for convergence considered in Art. 43 may be restated for the class of functions now under consideration. The proofs are very similar to those already considered and may be left for the reader to supply. These tests are given in the following theorems.

THEOREM I. *Let $F(x)$ be positive for all values of $x$ in the interval $(a, b)$ and suppose the integral $\int_a^b F(x)dx$ converges. If the given function $f(x)$ satisfies the condition that $|f(x)| \leqq F(x)$ for all values of $x$ in the deleted neighborhoods of the points of infinite discontinuity of $f(x)$, then the integral $\int_a^b f(x)dx$ also converges.*

**Ex. 3.** Test the convergence of the integral $\int_0^1 \log x\, dx$.

The integrand has an infinite discontinuity at $x = 0$. We know that $\underset{x \to 0}{L}\, x^{\frac{1}{2}} \log x$ is zero. Hence, we may write $x^{\frac{1}{2}} \log x < a$, where $a$ is any finite number, or what is the same thing, we may write

$$|\log x| < \frac{a}{|x^{\frac{1}{2}}|}.$$

The integral $\int_0^1 \frac{a\,dx}{|x^{\frac{1}{2}}|}$ is convergent. Hence, by the foregoing theorem, the integral $\int_0 \log x\, dx$ is also convergent.

THEOREM II. *Let $F(x)$ be positive for all values of $x$ in the interval $(a, b)$ and suppose that the integral $\int_a^b F(x)dx$ is divergent for $x = a$. If $|f(x)| \geqq F(x)$ for all values of $x$ in the deleted neighborhood of $x = a$, then the integral $\int_a^b f(x)dx$ is divergent.*

**Ex. 4.** Show that the integral $\int_0^1 \frac{\sec x\,dx}{x}$ is divergent.

Since $|\sec x| \geqq 1$ for all values of $x$, we have

$$\left|\frac{\sec x}{x}\right| \geqq \frac{1}{x},$$

and the integral $\int_0^1 \frac{dx}{x}$ is known to diverge. Hence, the given integral is divergent.

THEOREM III. *If $f(x)$ is monotone and bounded, then $\int_a^b f(x)\phi(x)dx$ is convergent if $\int_a^b \phi(x)dx$ converges.*

**Ex. 5.** Test the convergence of the integral $\int_0^1 e^{-\frac{1}{x^2}} \log x\,dx$.

Since the integral $\int_0^1 \log x\,dx$ converges and in the given interval $e^{-\frac{1}{x^2}}$ is bounded and monotone, it follows from the theorem that the given integral must also converge.

It must not be assumed that because the integral of each of two functions converges the integral of their product converges. For example, if we put

$$f(x) = \phi(x) = \frac{1}{x^{\frac{1}{2}}},$$

then the two integrals

$$\int_0^1 f(x)dx = \int_0^1 \frac{dx}{x^{\frac{1}{2}}}, \quad \int_0^1 \phi(x)dx = \int_0^1 \frac{dx}{x^{\frac{1}{2}}},$$

both converge. However, the integral

$$\int_0^1 f(x)\cdot\phi(x)dx = \int_0^1 \frac{dx}{x},$$

does not converge.

THEOREM IV. *If there exists a number $\mu$, where $0 < \mu < 1$, such that for all values of $x$ in the deleted neighborhood of $a$ we have*

$$(x - a)^{\mu}\,|f(x)| < M,$$

*where $M$ is any finite number, then the integral $\int_a^b f(x)dx$ converges.*

**Ex. 6.** Test the convergence of the integral $\int^1 \frac{\sin\frac{1}{x}}{x^{\frac{1}{2}}}\,dx$.

We have for $\mu = \frac{1}{2}$

$$x^{\frac{1}{2}}\,|f(x)| = \left|\sin\frac{1}{x}\right| < 1.$$

Consequently, the integral $\int_0^1 \frac{\sin\frac{1}{x}}{x^{\frac{1}{2}}}\,dx$ must converge for $x = 0$.

THEOREM V. *The integral $\int_a^b f(x)dx$ is convergent if the integral $\int_a^b |f(x)|\,dx$ is convergent.*

**Ex. 7.** Consider the convergence of the integral $\int_0^1 \frac{\cos\frac{1}{x}}{\sqrt{1-x^2}}\,dx.$

According to the foregoing theorem, this integral converges if the integral $\int_0^1 \frac{\left|\cos\frac{1}{x}\right| dx}{\sqrt{1-x^2}}$ converges. We have

$$\int_0^1 \frac{\left|\cos\frac{1}{x}\right| dx}{\sqrt{1-x^2}} \leqq \int_0^1 \frac{dx}{\sqrt{1-x^2}} = \text{arc sin } x\Big]_0^1 = \frac{\pi}{2}.$$

As this last integral exists, the integral $\int_0^1 \frac{\left|\cos\frac{1}{x}\right| dx}{\sqrt{1-x^2}}$ must converge, and hence the given integral converges.

**45. Properties of infinite integrals.** The laws of operation with definite integrals were established for bounded functions defined over a finite interval. We need to inquire how far they may be extended to infinite integrals. It is helpful to keep in mind that, while a finite definite integral is the limit of a summation of an increasing number of small elements, the infinite integral is the limit of another integral. It follows at once, from the definition of an infinite integral, that certain laws of operation with finite integrals hold likewise for infinite integrals. For example, the integral of a constant times a function is that constant times the integral of the function. Likewise the sum of two integrals is the integral of the sum of the functions.

The converse, however, is not necessarily true; that is, in dealing with infinite integrals we can not always say that

$$\int_a^b [f_1(x) + f_2(x)]\,dx = \int_a^b f_1(x)dx + \int_a^b f_2(x)dx. \tag{1}$$

For, it follows from the laws of limits that at least two of the three limits defining the integrals involved must exist in order to assert that the remaining limit also exists. For example, if it is known that the integral on the left converges, then if one of the integrals on the right converges the other does also; or if both integrals on the right converge, then the integral of the sum converges. However, if neither of the integrals on the right converges it is still possible for the integral on the left to converge to a definite limit. As a simple illustration, consider the two integrals

$$\int_0^{\frac{\pi}{2}} \frac{x \sin x - 1}{x}\,dx, \int_0^{\frac{\pi}{2}} \frac{dx}{x}.$$

Neither of these integrals converges because of the nature of the discontinuity at $x = 0$, but if we take the integral of the sum of the integrands, we have

$$\int_0^{\frac{\pi}{2}}\left[\frac{x \sin x - 1}{x} + \frac{1}{x}\right] dx = \int_0^{\frac{\pi}{2}} \sin x \, dx = 1.$$

A similar remark can be made for the application of the method of integration by parts, namely the application of the formula:

$$\int_a^b u \, dv = u \cdot v\Big]_a^b - \int_a^b v \, du, \tag{2}$$

where one of the limits of integration is infinite or the integrand has an infinite discontinuity. We can say, however, that if two of the three integrals converge, then the third integral also converges. Thus if each of the integrals involved in the second member of (2) converges, then the integral in the first member converges.

For example, consider the integral $\int_0^{\frac{2}{\pi}} \frac{1}{x} \sin \frac{1}{x} \, dx$. The integrand has an infinite discontinuity at $x = 0$. We have

$$\int_0^{\frac{2}{\pi}} \frac{1}{x} \sin \frac{1}{x} \, dx = \int_0^{\frac{2}{\pi}} x \sin \frac{1}{x} \cdot \frac{dx}{x^2}.$$

Put $$dv = \sin \frac{1}{x} \, \frac{dx}{x^2}, \quad u = x,$$

and thus obtain $$v = \cos \frac{1}{x}, \quad du = dx.$$

Hence from (2) we have

$$\int_\epsilon^{\frac{2}{\pi}} \frac{1}{x} \sin \frac{1}{x} \, dx = -\epsilon \cos \frac{1}{\epsilon} - \int_\epsilon^{\frac{2}{\pi}} \cos \frac{1}{x} \, dx.$$

The limit of $\epsilon \cdot \cos \frac{1}{\epsilon}$ as $\epsilon$ approaches zero is zero and the integral $\int_0^{\frac{2}{\pi}} \cos \frac{1}{x} \, dx$ exists, since $\cos \frac{1}{x}$ has but one point of discontinuity and is bounded. Consequently, the limit $\underset{\epsilon \to 0}{L} \int_\epsilon^{\frac{2}{\pi}} \cos \frac{1}{x} \, dx$ exists and the given integral must converge.

We shall now inquire whether the law for the change of variables, as stated in Art. 40, holds for infinite integrals. We shall consider the case where one of the limits of integration of the transformed integral becomes infinite. The case where the integrand has an infinite discontinuity may be stated in a similar manner. We shall assume that $\phi(t)$ is monotone, continuous and has a definite limit as $t$ becomes infinite. Let the interval $(a, b)$ for $x$ correspond to $(\alpha, \infty)$ for $t$. We have from Theorem XIII of Art. 40

$$\int_a^{b-\epsilon} f(x)dx = \int_\alpha^{t'} f[\phi(t)]\, D\phi(t)\, dt,$$

which holds for arbitrarily small values of $\epsilon$ and for all corresponding values of $t'$. Upon passing to the limit as $\epsilon$ approaches zero and therefore as $t'$ becomes infinite, the limit in the left-hand member exists and hence the one in the right-hand member must also exist. That is, the law for change of variable applies to infinite integrals as for finite integrals. In case the integrand has an infinite discontinuity the change of variable should be examined for each subinterval within which the integral is bounded and the limit be then considered just as was done for the convergence of the integral.

It follows at once from the definition of an infinite integral that if $f(x)$ is integrable in $(a, b)$, or if the integral $\int_a^\infty f(x)dx$ exists, then it is integrable in any subinterval, and we may write

$$\int_a^b (x)dx = \int_a^{x_0} f(x)dx + \int_{x_0}^b f(x)dx,\ a < x_0 < b,$$

$$\int_a^\infty f(x)dx = \int_a^{x_0} f(x)dx + \int_{x_0}^\infty f(x)dx,\ x_0 > a.$$

We have seen that the definition of an infinite integral can be extended to the case where the infinite discontinuities form a reducible set. It follows from this definition that if $f(x)$ is any integrable function having infinite discontinuities, the indefinite integral

$$\phi(x) \equiv \int_a^x f(x)dx,\ a \leqq x \leqq b,$$

is a continuous function of $x$, just as was the case for finite integrals.

**46. Integrals involving a parameter.** We shall now consider certain problems concerning the integral $\int_a^b f(x, y)dx$, where $y$ is to be regarded as a parameter. Suppose $f(x, y)$ to be defined for all values of $x$, $y$ in the rectangle $R$ given by the inequalities

$$a \leqq x \leqq b,\ y_0 < y \leqq d. \tag{1}$$

For the value $y = y_0$, let the given function be defined by the relation

$$f(x, y_0) = \underset{y \to y_0}{L} f(x, y_0).$$

For each value of $y \neq y_0$ we shall assume that $f(x, y)$ is an integrable function of $x$. We shall study the properties of the function defined by the integral of the given function, namely

$$\phi(x, y) \equiv \int_a^x f(x, y)dx, \; a \leqq x \leqq b.$$

We shall also inquire as to the conditions under which the limiting function $f(x, y_0)$ may be integrated.

We have the following theorems.

THEOREM I. *If $f(x, y)$ is an integrable function of $x$ in a given region $R$ defined by the inequalities* (1) *and if $f(x, y)$ is continuous in $(x, y)$ together in this region, then*

$$\phi(x, y) \equiv \int_a^x f(x, y)dx, \; a \leqq x \leqq b,$$

*is continuous in $x$ and in $y$ within $R$.*

From the existence of the integral $\int_a^x f(x, y)dx$, it follows that the indefinite integral $\phi(x, y)$ is continuous in $x$ for each $y \neq y_0$. To show that it is also continuous in $y$ at any point $(\alpha, \beta)$ within $R$, that is for all values of $x$ and $y$ for which the inequalities (1) hold, we have

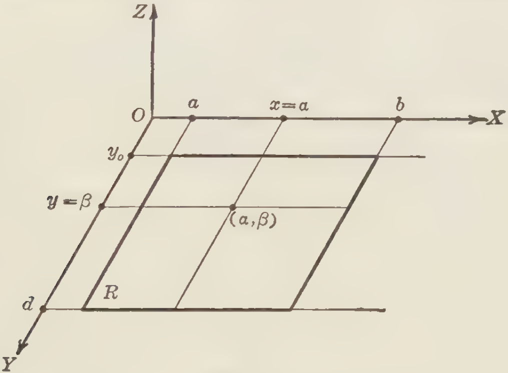


FIG. 35.

$$\phi(\alpha, \beta + \Delta y) - \phi(\alpha, \beta) = \int_a^\alpha f(x, \beta + \Delta y)dx - \int_a^\alpha f(x, \beta)dx$$

$$= \int_a^\alpha [f(x, \beta + \Delta y) - f(x, \beta)]dx.$$

Since $f(x,y)$ is continuous in $(x, y)$ together in the given region, it follows that $f(x,y)$ converges uniformly as $y$ approaches $\beta$, and hence we have, for an arbitrary value of $\epsilon$,

$$| f(x, \beta + \Delta y) - f(x, \beta) | < \epsilon, | \Delta y | < \delta,$$

and obtain

$$| \phi(\alpha, \beta + \Delta y) - \phi(\alpha, \beta) | < \epsilon \left| \int_a^\alpha dx \right| = \epsilon(\alpha - a) = \epsilon_1.$$

As $\epsilon$ is arbitrarily small and $(\alpha - a)$ is finite, it follows that $\epsilon_1$ is also arbitrarily small and hence that $\phi(x, y)$ is continuous in $y$ at $(\alpha, \beta)$. But $(\alpha, \beta)$ is any point within $R$. Hence, $\phi(x, y)$ is continuous in $y$ everywhere within this region.

COR. *If $f(x, y)$ is for each value of $y$ an integrable function of $x$ within a given region $R$, then within this region the integral*

$$\int_a^b f(x, y)dx \equiv F(y)$$

*defines a continuous function $F(y)$, provided $f(x, y)$ is continuous in $(x, y)$ together.*

The corollary follows as an immediate consequence of the foregoing theorem, and the details of the proof are left to the reader to supply.

THEOREM II. *Given a function $f(x, y)$ which is bounded and integrable as to $x$ in the interval $(a, b)$ for each value of $y \neq y_0$, and let $f(x, y_0)$ be bounded and defined by the limit*

$$f(x, y_0) = \underset{y \to y_0}{L} f(x, y).$$

*A necessary and sufficient condition that $f(x, y_0)$ is integrable in $(a, b)$ is that $f(x, y)$ converges quasi-uniformly in general to $f(x, y_0)$ in the given interval.*[1]

For each value of $y \neq y_0$ the function $f(x, y)$ is by hypothesis a bounded, integrable function of $x$ in $(a, b)$. Consequently, for any value of $y \neq y_0$ the points of discontinuity of $f(x, y)$ in $x$ form at most a set of measure zero. Any point of discontinuity in $x$ for one value of $y$ may or may not be a point of discontinuity of $f(x, y)$ for another value of $y$.

Since $f(x, y)$ converges quasi-uniformly in general, there are certain subintervals of $(a, b)$ in which the convergence is quasi-uniform.

---

[1] See Arzelà, *Sulle serie di funsioni,* Part 2 (1900), pp. 8–27.

For such a subinterval, this condition requires that there exist a set of values $\{y_n\}$ dense at $y_0$ and on the lines $y = y_n$ a finite set of intervals satisfying certain conditions. Denote by $E_n$ the points on $y = y_n$ where the function $f(x, y_n)$ is discontinuous. The measure of $E_n$ is zero, since by hypothesis $f(x, y_n)$ is an integrable function of $x$. Consider the sequence of point sets $\{E_n\}$ corresponding to the sequence of values $\{y_n\}$. Let $E$ be the limiting set of this sequence of point sets. Since $E_n$ has the measure zero, we have from Theorem X, Art. 21,

$$\underset{n\to\infty}{L} m(E_n) = 0 = m(E).$$

As $E$ is the limiting set of $\{E_n\}$, it includes those points where $f(x, y_0)$ is discontinuous. The points of discontinuity of $f(x, y_0)$ which occur in those subintervals of $(a, b)$ where $f(x, y)$ converges quasi-uniformly can then be inclosed in a set of intervals the sum of whose lengths $I_1$ can be made as small as we please.

To prove the given condition to be sufficient, we assume that $f(x, y)$ converges quasi-uniformly in general to the bounded limiting function $f(x, y_0)$. The sum $I_2$ of the lengths of the intervals where $f(x, y)$ fails to converge quasi-uniformly is by hypothesis arbitrarily small. In the remaining portion of the given interval, namely

$$(a, b) - (I_1 + I_2),$$

the limiting function $f(x, y_0)$ is continuous; that is, $f(x, y_0)$ is continuous in $(a, b)$ except at most for a set of points of measure zero. It is also bounded by the conditions of the theorem. Consequently, $f(x,y_0)$ is integrable in $(a, b)$ and the condition given in the theorem is sufficient.

The given condition is also necessary. For, let us assume that $f(x,y_0)$ is bounded and integrable in $(a, b)$. The points of $(a, b)$ where the oscillation of $f(x, y_0)$ is equal to or exceeds an arbitrary number $\frac{\eta}{4}$ form a set of content zero and hence is one which can be inclosed in a finite set of intervals the sum of whose lengths can be made arbitrarily small. Denote these intervals by $\delta_1^0, \delta_2^0, \cdots, \delta_p^0$. In the complementary set of intervals $C(\delta^0)$ the oscillation of $f(x,y_0)$ is less than $\frac{\eta}{4}$ at each point and there is associated with each value of $x$ a maximum interval $(x - \delta, x + \delta)$ for which we have

$$|f(x + h, y_0) - f(x, y_0)| < \frac{\eta}{4}, \quad |h| < \delta. \tag{2}$$

However, $\delta$ varies with $x$ and the aggregate of values which $\delta(x)$ may take has a greatest lower bound $\Delta_0$ greater than zero. For, let this lower bound be associated with the point $x_0$. But for $x_0$ we have an interval $[x_0 - \delta(x_0), x_0 + \delta(x_0)]$, where $\delta(x_0)$ is greater than zero. Then, in the interval $\left(x_0 - \frac{\delta(x_0)}{2}, x_0 + \frac{(\delta x_0)}{2}\right)$ there is associated with each $x$ within it a $\delta$ which is at least equal to $\frac{\delta(x_0)}{2}$. We can now associate with each point $x$ of $C(\delta^0)$ an interval $(x - \Delta_0, x + \Delta_0)$ for which (2) is valid, provided $|h| \leq \Delta_0$. It is understood that if $x$ is an end-point of one of the intervals $C(\delta^0)$ then only that part of the interval $(x - \Delta_0, x + \Delta_0)$ is to be retained which lies within $C(\delta^0)$.

Likewise, $f(x, y_n)$ is integrable for each $y_n \neq y_0$ and there exists a finite number of intervals $\delta_1^{(n)}, \delta_2^{(n)}, \cdots, \delta_q^{(n)}$, the sum of whose lengths is arbitrarily small, containing all values of $x$ where the oscillation of $f(x, y_n)$ is equal to or exceeds $\frac{\eta}{4}$. For the complementary set of intervals, there exists a $\Delta_n$ greater than zero such that for each $x$ we have

$$|f(x + h, y_n) - f(x, y_n)| < \frac{\eta}{4}, \; |h| < \Delta_n. \tag{3}$$

Denote by $\Delta_{0n}$ the smaller of the two values $\Delta_0, \Delta_n$. After omitting from $(a, b)$ the intervals $\delta^0, \delta^{(n)}$, we may associate with each $x$ in the remaining intervals an interval $(x - \Delta_{0n}, x + \Delta_{0n})$ for which (2) and (3) are both valid. As the selection of $y_n$ is arbitrary, there may exist values of $x$ in the intervals $(x - \Delta_{0n}, x + \Delta_{0n})$ for which

$$|f(x, y_n) - f(x, y_0)| > \eta. \tag{4}$$

By combining (2), (3), (4), we have

$$|f(x + h, y_n) - f(x + h, y_0)| > \frac{\eta}{2}, \; x - \Delta_{0n} < x + h < x + \Delta_{0n}. \tag{5}$$

Consider now any value of $x$ for which (4) is valid. Either the two functions $f(x, y_n)$, $f(x, y_0)$ have each an oscillation less than $\frac{\eta}{4}$ at $x$, or at this point at least one of them has an oscillation greater than $\frac{\eta}{4}$. In the first case, the point $x$ belongs to one of the intervals $\delta_1', \delta_2', \cdots$ for which (5) is valid, and in the second case it belongs to the set of intervals $\delta^0, \delta^{(n)}$ already excluded and need not be further considered. Let $(\alpha_1, \beta_1)$ be any interval on $y_n$ for which (5) holds. Because of the existence of the limit

$$\underset{y_n \to y_0}{L} f(x, y_n) = f(x, y_0),$$

there exists a value $y_{n_1}$, such that for all values of $y_n$ between $y_{n_1}$ and $y_0$, we have for the given value of $x$

$$|f(x, y_n) - f(x, y_0)| < \frac{\eta}{2}.$$

Combined with (2) and (3), we have

$$|f(x+h, y_n) - f(x+h, y_0)| < \eta, \ \ y_{n_1} < y_n < y_0.$$

From this relation it follows that as $y_n$ approaches $y_0$ the length of the segment $(\alpha_1, \beta_1)$ for which (5) holds must approach zero.

Having chosen a value $\eta$ and a particular $y_n$, we shall now consider the aggregate of points $x$ for which (5) is valid and show that the number of intervals $\delta'$ is finite and the sum of their lengths is arbitrarily small. Denote by

$$(a_1, b_1), (a_2, b_2), \cdots, (a_k, b_k), \cdots, (a_n, b_n)$$

the closed intervals after the subintervals $\delta^0$, $\delta^{(n)}$ have been omitted from $(a, b)$. Let $x$ be a point in one of the intervals $(a_k, b_k)$. There exists between $y_n$ and $y_0$ one or more values of $y_{n+p}$ for which

$$|f(x, y_0) - f(x, y_{n+p})| < \frac{3}{2}\eta, \tag{6}$$

while at the same time the $x$-oscillation is less than $\frac{\eta}{4}$. About each $x$ there exists an interval for which (6) holds, which interval lies on one side if $x$ is an extreme point of the interval. This interval varies in length with $y_{n+p}$ and as a function of $y_{n+p}$ has a least upper bound $L$. The value of $L$ varies with $x$ and the values which it may take have a greatest lower bound $l$, greater than zero, since the intervals $(a_k, b_k)$ may be regarded as closed and $L$ is positive.

Now begin with $a_1$. On some line $y = y_{n+p}$ there exists an interval $\delta_{a_1}$ to the right of $a_1$ for every point of which (6) is valid, where the length of $\delta_{a_1} \geqq l$. Next take the point $a_1' \equiv a_1 + \delta_{a_1}$. For this point there will exist on some line $y = y_{n+p_2}$ an interval $\delta_{a_1'}$ whose length exceeds or at least is equal to $l$ and lying to the right of $a_1'$ for each point of which (6) is valid. Since $l$ is greater than zero, we may by continuing this process reach the point $b_1$ after a finite number of steps. In the same manner, we may cover all of the intervals $(a_k, b_k)$. Hence, the number of corresponding intervals $\delta'$ is also finite, and by taking $y_{n+p}$ sufficiently close to $y_0$, each of these intervals may be made arbitrarily small, and hence the sum of their lengths also arbitrarily small. As we have seen the sum of the lengths of the intervals $\Delta_{0n}$ is also arbitrarily small.

Consequently, $f(x, y_0)$ can be regarded as the limiting function toward which, except for a set of points that can be inclosed in a finite set of intervals the sum of whose lengths can be made arbitrarily small, $f(x, y)$ converges quasi-uniformly; that is, in $(a, b)$, $f(x, y)$ converges quasi-uniformly in general to $f(x, y_0)$.

THEOREM III. *Given a function* $f(x, y)$, *defined for the region* $a \leqq x \leqq b$, $c \leqq y \neq y_0 \leqq d$, *where we have*

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0), \tag{7}$$

*and suppose that* $f(x, y)$ *is an integrable function of* $x$ *for each* $y \neq y_0$ *and that there exists a finite number* $M$ *such that*

$$|f(x, y)| < M, \ a \leqq x \leqq b, \ y \neq y_0. \tag{8}$$

*We then have* $$\int_a^x f(x, y_0)dx = \underset{y \to y_0}{L} \int_a^x f(x, y)dx,$$

*if* $f(x, y)$ *converges to* $f(x, y_0)$ *quasi-uniformly in general in the interval* $(a, b)$.[1]

The function $f(x, y_0)$ is bounded and from Theorem II it follows at once that it is integrable in $(a, b)$. Moreover, as a consequence of (7) and (8) $f(x, y)$ is bounded when considered as a function of the two variables $(x, y)$. And in fact, we may write

$$|f(x, y_0) - f(x, y)| < 2M, \ a \leqq x \leqq b, \ c \leqq y \leqq d. \tag{9}$$

As both $f(x, y_0)$ and $f(x, y)$, $y \neq y_0$, are integrable in $x$, then

$$F(x, y) \equiv f(x, y_0) - f(x, y),$$

considered as a function of $x$, is integrable in any subinterval $(a, x)$ of $(a, b)$.

Denote by $\eta$, $\epsilon$ two independent, arbitrarily small positive numbers. From the conditions given the theorem, it follows that for $y$ sufficiently close to $y_0$ the points $x$ where

$$|f(x, y_0) - f(x, y)| \geqq \eta \tag{10}$$

can be inclosed in intervals the sum of whose lengths is less than $\epsilon$.

Let $x'$ be any one of the values of $x$ for which the inequality (10) is valid for a particular valve of $y$. We shall distinguish two cases. First, where $F(x, y)$ has at $x'$ an oscillation less than $\frac{\eta}{2}$ and second where the oscillation is equal to or greater than $\frac{\eta}{2}$. As $F(x, y)$ is integrable, its points of discontinuity where the oscillation is equal to or greater than $\frac{\eta}{2}$ must form a closed set of content zero. Consequently, we can inclose all such values of $x$ in intervals the sum of whose lengths is less than $\frac{\epsilon}{2}$. If, on the other hand, $F(x, y)$ has at $x'$ an oscillation less than $\frac{\eta}{2}$, then there must exist an interval, at each point $x$ of which we have

---

[1] See Arzelà, *Sulle serie di funzioni*, Part 2 (1900), § 7.

$$|f(x, y_0) - f(x, y)| > \frac{\eta}{2}.$$

As already shown in the proof of Theorem II, it follows that the sum of the lengths of these intervals has the limit zero as $y$ approaches $y_0$ and hence can be made less than $\frac{\epsilon}{2}$ for all values of $y$ sufficiently close to $y_0$.

It now follows that there exists a $y_n$ such that for all values of $y$ lying between $y_0$ and $y_n$ the points $x'$ lie in intervals the sum of whose lengths is less than $\epsilon$. Then for all values of $x$ in $(a, b)$ and all values of $y$ between $y_n$ and $y_0$ we have

$$\int_a^x \{f(x, y_0) - f(x, y)\}dx < \eta(x - a) + 2\,\epsilon M. \tag{11}$$

Since $\eta$ and $\epsilon$ are both arbitrarily small and $(x - a)$ and $M$ are finite, we have in the limit as $y$ approaches $y_0$

$$\int_a^x f(x, y_0)dx = \underset{y\to y_0}{L} \int_a^x f(x, y)dx,$$

which is the desired result.

THEOREM IV. *Given a function* $f(x, y)$ *defined for the region* $a \leqq x \leqq b,\ c \leqq y \neq y_0 \leqq d$, *where for* $y_0$ *we have*

$$\underset{y\to y_0}{L} f(x, y) = f(x, y_0),$$

*and where* $f(x, y)$ *is a bounded, integrable function of* $x$ *for each* $y \neq y_0$. *Suppose that* $f(x, y_0)$ *is bounded and that the limit* $\underset{y\to y}{L} \int_a^x f(x, y)\,dx$ *exists and defines a continuous function of* $x$. *Then*

$$\int_a^x f(x, y_0)dx = \underset{y\to y_0}{L} \int_a^x f(x, y)dx,\ a \leqq x \leqq b,$$

*provided the points* $(x, y_0)$ *in the neighborhood of which* $f(x, y)$ *has no finite upper bound in* $(x, y)$ *form at most an enumerable set and* $f(x, y)$ *converges quasi-uniformly to the limiting function* $f(x, y_0)$.[1]

---

[1] See Arzelà, *Sulle series di funzioni*, Part 2 (1900), §12.

As $f(x, y)$ converges quasi-uniformly to the function $f(x, y_0)$, it follows that $f(x, y_0)$ is an integrable function. From the hypothesis, there may exist on the line $y=y_0$ certain points $x'$ such that one can not assign an interval $(x' - \delta_1', x' + \delta_2')$ and a value $y = y_n$ so that for all values of $x$ within this interval and all values of $y$ between $y_n$ and $y_0$ we have

$$| f(x, y) | < M,$$

no matter how large $M$ may be chosen.

The points $x'$ form a closed set $E$. If $E$ is enumerable, then the derived set $E'$ is also enumerable, since it is a subset of $E$. In this case, $E$ has the content zero by Theorem III, Art. 20.

From the integrability of $f(x, y_0)$, it follows that

$$\phi(x, y_0) \equiv \int_a^x f(x, y_0)dx$$

is a continuous function. For convenience, put

$$\psi(x, y) \equiv \int_a^x f(x, y)dx.$$

By hypothesis, the function $\psi(x, y_0)$, where

$$\psi(x, y_0) \equiv \underset{y \to y_0}{L}\, \psi(x, y) = \underset{y \to y_0}{L} \int_a^x f(x, y)dx,$$

is also continuous. We shall show that

$$\phi(x, y_0) = \psi(x, y_0).$$

Let $x$ be a point of the set $C(E)$. We can then assign to $x$ an interval $(x-\delta_1, x + \delta_2)$ and a value $y = y_n$ such that for any value of $x$ in this interval and any $y$ between $y_n$ and $y_0$, we have

$$| f(x, y) | < M,$$

where $M$ is a finite number. For these values of $x$, let us consider the difference ratios

$$\frac{\phi(x+\Delta x, y_0)-\phi(x, y_0)}{\Delta x}, \quad \frac{\psi(x+\Delta x, y_0)-\psi(x, y_0)}{\Delta x},$$

where $x + \Delta x$ is so chosen as to lie in the interval $(x - \delta_1, x + \delta_2)$. Let $\lambda$ be an arbitrarily small positive number. Since for each $x$ in $C(E)$ we have

$$\psi(x, y_0) = \underset{y \to y_0}{L}\, \psi(x, y),$$

we can so choose a $y_n$ that for each $y$ between $y_n$ and $y_0$ we have

$$\frac{\psi(x + \Delta x, y_0) - \psi(x, y_0)}{\Delta x} \quad \frac{\psi(x + \Delta x, y) - \psi(x, y)}{\Delta x} + \mu\lambda,$$

where $-1 < \mu < 1$. From the definition of $\psi(x, y)$, we have

$$\frac{\psi(x + \Delta x, y) - \psi(x, y)}{\Delta x} = \frac{1}{\Delta x}\int^{x+\Delta x} f(x, y)dx,$$

where $x$ and $x + \Delta x$ lie in the interval $(x-\delta_1, x+\delta_2)$. From Theorem III, the integral $\int_x^{x+\Delta x} f(x, y)dx$ differs from $\int_x^{x+\Delta x} f(x, y_0)dx$ by less than $\lambda$ for each $y$ sufficiently close to $y_0$. Hence, we have

$$\frac{1}{\Delta x}\int_x^{x+\Delta x} f(x, y)dx = \frac{1}{\Delta x}\int_x^{x+\Delta x} f(x, y_0)dx + \mu'\frac{\lambda}{\Delta x},$$

where $-1 < \mu' < 1$, and $\lambda$ is independent of $x$ and $\Delta x$. From the foregoing relations, we have

$$\begin{aligned}\frac{\psi(x + \Delta x, y_0) - \psi(x, y_0)}{\Delta x} &= \frac{1}{\Delta x}\int_x^{x+\Delta x} f(x, y_0)dx + \mu\lambda + \mu'\frac{\lambda}{\Delta x}\\ &= \frac{\phi(x + \Delta x, y_0) - \phi(x, y_0)}{\Delta x} + \lambda\left(\mu + \frac{\mu'}{\Delta x}\right).\end{aligned}$$

As this relation is valid for every pair of values $x, x + \Delta x$ chosen in the manner indicated, it follows that for each point $x$ of $C(E)$, we have in the limit as $\lambda$ approaches zero

$$\frac{\psi(x + \Delta x, y_0) - \psi(x, y_0)}{\Delta x} = \frac{\phi(x + \Delta x, y_0) - \phi(x, y_0)}{\Delta x}.$$

As $\Delta x$ approaches zero, these ratios give us two corresponding derivative numbers of $\psi(x, y_0)$, $\phi(x, y_0)$, say $D^+\psi(x, y_0)$, $D^+\phi(x, y_0)$.

By hypothesis, $E$ is an enumerable set; hence, it follows (Art. 31) that for all values of $x$ in $(a, b)$ we have

$$\psi(x, y_0) = \phi(x, y_0) + c,$$

where $c$ is a constant. However, $c = 0$, since for $x = a$, we have

$$\psi(a, y_0) = \phi(a, y_0) = 0.$$

Consequently, we have for all values of $x$ in $(a, b)$

$$\psi(x, y_0) = \phi(x, y_0);$$

that is, $$\int_a^x f(x, y_0)dx = \underset{y \to y_0}{L} \int_a^x f(x, y)dx, \quad a \leqq x \leqq b,$$

which establishes the theorem.

As uniform convergence is a special case of quasi-uniform convergence, we have also as a corollary to the foregoing theorem the following:

COR. *Let* $f(x, y)$ *be a bounded, integrable function of* $x$ *for* $a \leqq x \leqq b$, $c \leqq y \neq y_0 \leqq d$, where

$$\underset{y \to y_0}{L} f(x, y) = f(x, y_0),$$

*and* $f(x, y_0)$ *is bounded. Then, if* $f(x, y)$ *converges uniformly to* $f(x, y_0)$ *and if* $\underset{y \to y_0}{L} \int_a^x f(x, y)dx$ *defines a continuous function, we have*

$$\int_a^x f(x, y_0) = \underset{y \to y_0}{L} \int_a^x f(x, y)dx.$$

If we replace the parameter $y$ by the subscript $n$, we may apply the foregoing theorems to the integration of a sequence of functions $\{f_n(x)\}$. Let us suppose that for each value of $n$ the function $f_n(x)$ is bounded and integrable in the interval $(a, b)$, and let $\{f_n(x)\}$ converge to the bounded function $f(x)$. It then follows from Theorem II that the necessary and sufficient condition that $f(x)$ be integrable is that the given sequence $\{f_n(x)\}$ shall converge quasi-uniformly in general in the interval $(a, b)$. If in addition to quasi-uniform convergence in general we have the condition that for all values of $n$ and of $x$ in $(a, b)$

$$| f_n(x) | < M,$$

where $M$ is some finite number, or if the convergence of $\{f_n(x)\}$ is uniform to a bounded function, it then follows from Theorem III that

$$\int_a^x f(x)dx = \underset{n \to \infty}{L} \int_a^x f_n(x)dx, \; a \leqq x \leqq b. \tag{12}$$

If, on the other hand, we have the condition that $f_n(x)$ is integrable and bounded for each value of $n$ and the limit $\underset{n \to \infty}{L} \int_a^x f_n(x)dx$ exists and is a continuous function, but that as $n$ increases indefinitely $f_n(x)$ exceeds all finite bounds in the neighborhood of certain points of $(a, b)$, then we have from Theorem IV the relation expressed in (12), provided these points form an enumerable set $E$, and $\{f_n(x)\}$ converges quasi-uniformly. From the foregoing corollary, it follows that it is sufficient if the convergence is uniform.

In Theorem I, we considered the continuity of $F(y)$, where

$$F(y) \equiv \int^b f(x, y)dx$$

and $a, b$ are constants. We shall now consider the case where $a$ and $b$ are replaced by functions of $y$, say $\phi_1(y)$ and $\phi_2(y)$, respectively. We may state the following theorem.

THEOREM V. *Let $F(y)$ be defined by the relation*

$$F(y) = \int_{\phi_1(y)}^{\phi_2(y)} f(x, y)dx, \ a \leqq x \leqq b, \ c \leqq y \leqq d ,$$

*where $\phi_1(y)$, $\phi_2(y)$ are continuous and $f(x, y)$ is continuous in $(x, y)$ together. Then $F(y)$ is also continuous.*

From the conditions of the theorem we may write

$$\begin{aligned} F(y + \Delta y) - F(y) & \\ &= \int_{\phi_1(y+\Delta y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx - \int_{\phi_1(y)}^{\varphi_2(y)} f(x, y)dx \\ &= \int_{\phi_1(y)}^{\phi_2(y)} [f(x, y + \Delta y) - f(x, y)]dx \\ &\quad - \int_{\phi_1(y)}^{\phi_1(y+\Delta y)} f(x, y + \Delta y)dx + \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx \\ &= \int_{\phi_1(y)}^{\phi_2(y)} [f(x, y + \Delta y) - f(x, y)]dx \\ &\quad + \int_{\phi_1(y+\Delta y)}^{\phi_1(y)} f(x, y + \Delta y)dx + \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx. \end{aligned}$$

Taking absolute values, we have

$$\begin{aligned} | F(y + \Delta y) - F(y) | \leqq & \left| \int_{\phi_1(y)}^{\phi_2(y)} [f(x, y + \Delta y) - f(x, y)]dx \right| \\ & + \left| \int_{\phi_1(y+\Delta y)}^{\phi_1(y)} f(x, y + \Delta y)dx \right| + \left| \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx \right| . \quad (13) \end{aligned}$$

Because of the continuity of $f(x, y)$, we may put

$$| f(x, y + \Delta y) - f(x, y) | < \epsilon,$$

whence the second member of (13) may be written, if we make use of Theorem X, Art. 40,

$$\begin{aligned} | F(y + \Delta y) - F(y) | < \epsilon \int_{\phi_1(y)}^{\phi_2(y)} | dx | + | f(x_0, y + \Delta y) | \left| \int_{\phi_1(y+\Delta y)}^{\phi_1(y)} \right| dx | \\ + | f(x_1, y + \Delta y) | \left| \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} \right| dx |, \end{aligned}$$

where $x_0$ is some value between $\phi_1(y)$ and $\phi_1(y + \Delta y)$ and $x_1$ lies between $\phi_2(y)$ and $\phi_2(y + \Delta y)$. This inequality gives upon integration

$$\begin{aligned} | F(y + \Delta y) - F(y) | < \epsilon | \phi_2(y) - \phi_1(y) | + | f(x_0, y + \Delta y) | \Delta\phi_1 | \\ + | f(x_1, y + \Delta y) | \Delta\phi_2 | . \end{aligned}$$

By taking $\Delta y$ sufficiently small we can make $\epsilon$, $\Delta\phi_1$, $\Delta\phi_2$ arbitrarily small and hence obtain the relation

$$|F(y + \Delta y) - F(y)| < \epsilon_1.$$

It follows that $F(y)$ is continuous, as the theorem states.

**Ex.** Discuss the continuity of $F(y)$ defined by the integral

$$\int_y^{y^2} \frac{dx}{y^2 + x^2},\ y > 0.$$

Here both limits of integration are continuous functions of $y$, and the integrand is continuous in $x$ and $y$ together. Hence, the conditions of the theorem are satisfied and the integral defines a continuous function of $y$.

THEOREM VI. *Given the function $F(y)$ defined by the relation*

$$F(y) = \int_a^b f(x, y)dx,\ a \leqq x \leqq b,\ c \leqq y \leqq d.$$

*The derivative of $F(y)$ is given by the relation*

$$\frac{dF}{dy} = \int_a^b \frac{\partial f}{\partial y} dx,$$

*provided the function $f(x, y)$ is continuous in $(x, y)$ and has a partial derivative which is also continuous in $(x, y)$.*

We have for any value $y_0$ of $y$

$$\begin{aligned}\frac{F(y_0 + \Delta y) - F(y_0)}{\Delta y} &= \frac{1}{\Delta y}\int_a^b [f(x, y_0 + \Delta y) - f(x, y_0)]\, dx \\ &= \int_a^b \left[\frac{f(x, y_0 + \Delta y) - f(x, y_0)}{\Delta y}\right] dx. \qquad (14)\end{aligned}$$

However, by the Law of the Mean, we have

$$\begin{aligned}\frac{f(x, y_0 + \Delta y) - f(x, y_0)}{\Delta y} &= f_y'(x, \xi),\quad y_0 \leqq \xi \leqq y_0 + \Delta y, \\ &= f_y'(x, y_0) + \eta(\Delta y),\end{aligned}$$

where $\eta$ may be chosen independently of $x$, because of the continuity of $f_y'(x, y)$ in $(x, y)$. Moreover, we have

$$\underset{\Delta y \to 0}{L}\ \eta(\Delta y) = 0.$$

From (14) it follows that

$$\begin{aligned}\frac{F(y_0 + \Delta y) - F(y_0)}{\Delta y} &= \int_a^b f_y'(x, y_0)dx + \int_a^b \eta\, dx \\ &= \int_a^b f_y'(x, y_0)dx + \eta(b - a),\end{aligned}$$

whence upon passing to the limit we obtain the desired result

$$\left.\frac{dF}{dy}\right]_{y_0} = \int_a^b f_y'(x, y_0)dx$$

for any value $y_0$ of the variable $y$.

In case the limits of integration are functions of the parameter $y$, we have the following theorem.

THEOREM VIII. *Let $f(x, y)$ be such a function that $\frac{\partial f}{\partial y}$ exists and suppose $f(x, y)$ and $\frac{\partial f}{\partial y}$ to be continuous in $x$ and $y$ together, where $a \leqq x \leqq b$, $c \leqq y \leqq d$. Then if $\frac{d\phi_1}{dy}$, $\frac{d\phi_2}{dy}$ exist and we put*

$$F(y) \equiv \int_{\phi_1(y)}^{\phi_2(y)} f(x, y)dx,$$

*we may write*

$$\frac{dF}{dy} = \int_{\phi_1(y)}^{\phi_2(y)} \frac{\partial f}{\partial y}\, dx - f(\phi_1, y)\frac{d\phi_1}{dy} + f(\phi_2, y)\frac{d\phi_2}{dy}.$$

We have at once

$$\frac{F(y + \Delta y) - F(y)}{\Delta y} = \frac{1}{\Delta y}\left[\int_{\phi_1(y+\Delta y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx \right.$$
$$\left. - \int_{\phi_1(y)}^{\phi_2(y)} f(x,y)dx\right]. \qquad (15)$$

However, the first of these integrals may be written

$$\int_{\phi_1(y+\Delta y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx = \int_{\phi_1(y)}^{\phi_2(y)} f(x, y + \Delta y)dx$$
$$+ \int_{\phi_1(y+\Delta y)}^{\phi_1(y)} f(x, y + \Delta y)dx + \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} f(x, y + \Delta y)dx.$$

Substituting in (15), we obtain after combining terms

$$\frac{F(y + \Delta y) - F(y)}{\Delta y} = \int_{\phi_1(y)}^{\phi_2(y)} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}\, dx$$
$$+ \int_{\phi_1(y+\Delta y)}^{\phi_1(y)} \frac{f(x, y+\Delta y)dx}{\Delta y} + \int_{\phi_2(y)}^{\phi_2(y+\Delta y)} \frac{f(x, y+\Delta y)dx}{\Delta y}$$
$$= \int_{\phi_1(y)}^{\phi_2(y)} f_y'(x, y + \theta\Delta y)dx + \frac{f(x_0, y + \Delta y)}{\Delta y}\int_{\phi_1(y+\Delta y)}^{\phi_1(y)} dx$$
$$+ \frac{f(x_1, y + \Delta y)}{\Delta y}\int_{\phi_2(y)}^{\phi_2(y+\Delta y)} dx,$$

where $0 < \theta < 1$, and $x_0$, $x_1$ are points between $\phi_1(y)$, $\phi_1(y + \Delta y)$ and $\phi_2(y)$, $\phi_2(y + \Delta y)$, respectively, as required by the Law of the

Mean for differentiation and for integration. Evaluating the last two integrals, we have

$$\begin{aligned}\frac{F(y+\Delta y)-F(y)}{\Delta y} &= \int_{\phi_1(y)}^{\phi_2(y)} f'_y(x, y+\theta\Delta y)dx \\ &\quad - f(x_0, y+\Delta y)\,\frac{\phi_1(y+\Delta y)-\phi_1(y)}{\Delta y} \\ &\quad + f(x_1, y+\Delta y)\,\frac{\phi_2(y+\Delta y)-\phi_2(y)}{\Delta y}.\end{aligned}$$

Upon passing to the limit as $\Delta y$ approaches zero, remembering that $f'_y(x, y)$ is continuous in $(x, y)$ and that $x_0$ and $x_1$ become $\phi_1$, $\phi_2$, respectively, we have the required result, namely

$$\frac{dF}{dy} = \int_{\phi_1(y)}^{\phi_2(y)} \frac{\partial f}{\partial y}dx - f(\phi_1, y)\frac{d\phi_1}{dy} + f(\phi_2, y)\frac{d\phi_2}{dy}.$$

**47. Uniform convergence of an infinite integral.** We shall first consider infinite integrals of the type $\int_a^\infty f(x, y)dx$, where $y$ is to be regarded as an arbitrary parameter lying between two constant values, $c$ and $d$. We shall assume that for each value of $y$ the integral converges. It then follows that for each value of $y$, say $y = y_1$, there exists a corresponding value of $x$, say $x = x_1$, such that we have

$$\left|\int_{x_1}^\infty f(x, y_1)dx\right| < \epsilon, \tag{1}$$

where $x_1$ is the greatest lower bound of those values of $x$ for which this inequality holds, and $\epsilon$ is an arbitrarily small positive number. As $y$ varies, the corresponding value of the lower limit of integration may also change in order that the inequality (1) may hold for the same value of $\epsilon$. As $y$ takes all values in the interval $(c, d)$, $\epsilon$ remaining unchanged, the aggregate of the least values of $x$ which can be selected as the lower limit of integration may have a finite upper bound, which we shall denote by $x_0$. Then for a given $\epsilon$ this value $x = x_0$ is independent of the choice of $y$ and may be said

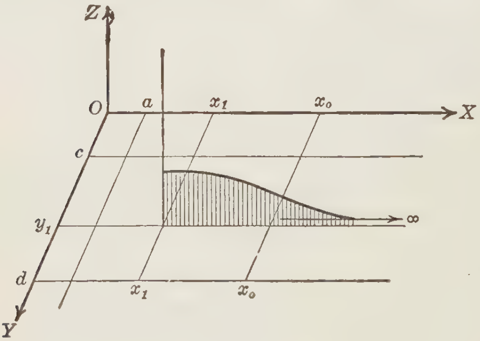


FIG. 36.

to answer uniformly well for all values of $y$. We say then that the given integral **converges uniformly** in the interval $(c, d)$.

Even if the given infinite integral does not converge uniformly, it is possible that for a given value of $x > a$, say $x_1$, there exists beyond $x_1$ a finite number of lines

$$x = x_s,\ s = s_1, s_2, \cdot \cdot \cdot, s_k,$$

such that for all values of $y$ in the interval $(c, d)$ and for a lower limit of integration taken on the broken line $x_s$ we have

$$\left| \int_{x_s}^{\infty} f(x, y)dx \right| < \epsilon,$$

where it is to be understood that the segments of this broken line project upon the line $x = a$ in such a manner that these projections exactly fill the interval $(c, d)$. If these conditions are satisfied for every choice of $x_1$ however large, then we say that the given infinite integral **converges quasi-uniformly** in the interval $(c, d)$. Of course any infinite integral which converges uniformly also converges quasi-uniformly, but the converse is not true.

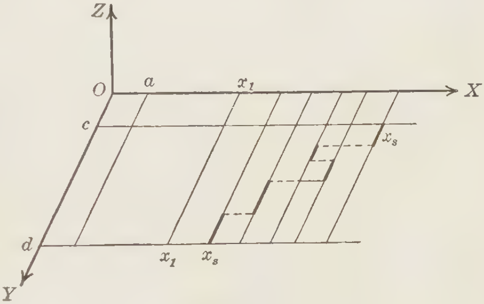


FIG. 37.

If the given integral converges quasi-uniformly in the given interval with the exception of a finite set of subintervals the sum of whose lengths can be made arbitrarily small, then we say that it **converges quasi-uniformly in general** in that interval.

In Art. 26 we considered certain tests for the uniform convergence of a function of two independent variables which are similar to those that may be stated for infinite integrals. Some of these tests are given in the following theorems.

THEOREM I. *For all values of $y$ in the interval $(c, d)$, let $f(x, y)$ satisfy the condition*

$$| f(x, y) | < M(x),$$

*where $M(x)$ is a finite, positive function of $x$ alone. Then if the integral $\int_a^{\infty} M(x)dx$ converges, the given integral $\int_a^{\infty} f(x, y)dx$ converges absolutely and uniformly in the interval $(c, d)$.*

Since $M(x)$ is independent of $y$ and the integral $\int_a^\infty M(x)dx$ converges, we can replace the lower limit of integration $a$ by a value $x_0$, which does not depend upon $y$, such that we have

$$\left| \int_{x_0}^\infty M(x)dx \right| < \epsilon.$$

We therefore have the relation

$$\left| \int_{x_0}^x f(x, y)dx \right| \leqq \int_{x_0}^x \left| f(x, y) \right| dx < \int_{x_0}^x M(x)dx < \int_{x_0}^\infty M(x)dx < \epsilon.$$

Hence, it follows that the given integral converges uniformly. That it also converges absolutely follows from the definition of absolute convergence.

Ex. 1. Test the infinite integral $\int_0^\infty \frac{\sin (xy)}{1 + x^2} dx$ for uniform convergence.

We have
$$\left| \frac{\sin (xy)}{1 + x^2} \right| < \frac{1}{1 + x^2} = M(x).$$

$$\int_0^x M(x)dx = \int_0^x \frac{dx}{1 + x^2} = \text{arctan } x.$$

The infinite integral $\int_0^\infty M(x)dx$ converges; for, we have

$$\int^\infty M(x)dx = \int_0^\infty \frac{dx}{1 + x^2} = \underset{x = \infty}{L} \text{ arctan } x = \frac{\pi}{2}.$$

The given function therefore satisfies the conditions of the foregoing theorem and hence converges both absolutely and uniformly in any finite interval for $y$.

THEOREM II. *If the integral* $\int_a^\infty f(x, y) \cdot \phi(x)dx$ *converges for each value of* $y$ *in the interval* $(c, d)$, *it then converges uniformly in this interval, provided the integral* $\int_a^\infty \phi(x)dx$ *converges absolutely and* $f(x, y)$ *is bounded in* $(x, y)$ *together.*

We have
$$| f(x, y) | \leqq M,$$
where $M$ is a constant. We may also write, since $\int_a^\infty \phi(x)dx$ converges absolutely,

$$\int_{x_1}^{x_2} | \phi(x) | dx < \frac{\epsilon}{M}, \quad x_0 < x_1 < x_2.$$

We thus obtain a value of $x_0$ independent of $y$ such that for $x_0 < x_1 < x_2$ we have

$$\left| \int_{x_1}^{x_2} f(x, y) \cdot \phi(x)dx \right| \leqq M \left| \int_{x_1}^{x_2} \phi(x)dx \right| \leqq M \int_{x_1}^{x_2} | \phi(x) | dx < \epsilon.$$

Consequently, the given integral converges uniformly in the interval $(c, d)$.

**Ex. 2.** Consider the uniform convergence of $\int_1^\infty \frac{\sin\frac{1}{x}}{x(x+y)^2}\,dx$ for values of $y > 1$.

We may let $\phi(x) = \frac{1}{x}\sin\frac{1}{x}$ and $f(x, y) = \frac{1}{(x+y)^2}$. It follows that $\int_1^\infty \phi(x)dx$ converges absolutely, since we have

$$\int_1^\infty \left|\frac{\sin\frac{1}{x}}{x}\right| dx \leqq \int_1^\infty \left|\frac{1}{x}\right| \cdot \left|\sin\frac{1}{x}dx,\right|$$

and the latter integral converges by Theorem III of Art. 43. Moreover, the function

$$f(x, y) = \frac{1}{(x+y)^2}$$

is bounded for all values of $x$ and $y$, where $x > 1$, $y > 1$. It then follows from the foregoing theorem that the given function converges uniformly for all finite values of $y > 1$.

We may now state the following theorem.

THEOREM III. *If $f(x, y)$ is continuous in $(x, y)$, $y_0 \leqq y \leqq d$, $x \geqq x_1$, a necessary and sufficient condition that the integral $\int_{x_1}^\infty f(x, y)dx$ defines a continuous function of $y$ is that this integral converges quasi-uniformly in the interval $(y_0, d)$.*

Put
$$F(x, y) \equiv \int_{x_1}^x f(x, y)dx, \quad y_0 \leqq y \leqq d, \; x_1 \leqq x.$$

Since $f(x, y)$ is a continuous function in $(x, y)$ together, it follows from Theorem I of Art. 46 that $\int_{x_1}^x f(x, y)dx$ defines a continuous function of $y$. It is also continuous in $x$, since $f(x, y)$ is integrable for every value of $y$ and for all finite values of the upper limit of integration. We also have the condition that the limit $\underset{x\to\infty}{L} F(x, y)$ exists for each value of $y$; that is, we may write

$$\underset{x\to\infty}{L} F(x, y_k) = F(\infty, y_k), \; y_0 \leqq y_k \leqq d.$$

By hypothesis $F(x, y)$ converges quasi-uniformly. Consequently, this function satisfies the conditions of Theorem XI of Art. 26, where the limiting value of the function is taken as one of the variables becomes infinite. It then follows that a necessary and sufficient condition that

$$F(\infty, y) \equiv \int_{x_1}^\infty f(x, y)dx$$

shall be continuous in $y$ is that given by the theorem.

As uniform convergence is a special case of quasi-uniform convergence, the following corollary follows at once from the foregoing theorem.

COR. *If* $f(x, y)$ *is a continuous function of* $(x, y)$ *in the region* $y_0 \leqq y \leqq d$, $x \geqq x_1$, *then the integral* $\int_{x_1}^{\infty} f(x, y)dx$ *defines a continuous function of* $y$ *if this integral converges uniformly.*

If the limiting function $F(\infty, y)$ is continuous, it is of course integrable. The above theorem, as also the corollary, gives a sufficient condition for integrability. A necessary and sufficient condition follows as in Theorem II of Art. 46 when we recall that the limit is now taken as $x$ becomes infinite instead of $y$ approaching a finite limit $y_0$. This condition can be stated as follows.

THEOREM IV. *If* $f(x, y)$ *is continuous in* $(x, y)$ *in the region* $y_0 \leqq y \leqq d$, $x \geqq x_1$, *a necessary and sufficient condition that* $F(y)$ *defined by the integral* $\int_{x_1}^{\infty} f(x, y)dx$ *is integrable in the given interval* $(y_0, d)$ *is that the given integral* $\int_{x_1}^{\infty} f(x, y)dx$ *converges quasi-uniformly in general in* $(y_0, d)$.

The discussions in this article have thus far concerned the type of infinite integrals where one or both of the limits of integration becomes infinite. Similar statements can be easily shown to hold for those infinite integrals where the integrand becomes infinite in the interval of integration.

**48. Double integrals. Repeated integrals.** Let $f(x, y)$ be a bounded function of the two independent variables $x$ and $y$ in the region $R$, defined by the inequalities

$$a \leqq x \leqq b, \quad c \leqq y \leqq d.$$

Denote by $\Delta_k(x, y)$ the rectangle formed by the increments $\Delta_k x$, $\Delta_k y$ of $x$ and $y$. Let $f(\xi_k, \eta_k)$ be the value of $f(x, y)$ at any point in $\Delta_k(x, y)$. As in the functions of a single variable, we form the sum

$$\sum_{k=1}^{n} f(\xi_k, \eta_k)\Delta_k(x, y)$$

taken over the region $R$. Let $\Delta(x, y)$ be the norm of the values $\Delta_k(x, y)$. If the limit

$$\underset{\Delta \to 0}{L} \sum_{k=1}^{n} f(\xi_k, \eta_k)\Delta_k(x, y) \tag{1}$$

exists as the norm $\Delta(x, y)$ approaches zero and simultaneously the number $n$ of rectangles increases indefinitely, we call this limiting value the **double integral** of $f(x, y)$ taken over the region $R$. Symbolically, we shall represent the double integral by

$$\int_R f(x, y)\, d(x, y).$$

If we replace $\Delta_k(x, y)$ by the product $\Delta_r x \cdot \Delta_s y$, we are led to an **iterated integral,** which we may denote by

$$\int_a^b dx \int_c^d f(x, y) dy.$$

If the order of integration is reversed, we have the iterated integral

$$\int_c^d dy \int_a^b f(x, y) dx.$$

An iterated integral is only successive integration of $f(x, y)$; first with respect to the one variable and then with respect to the other.

In a similar manner, we may define an integral of higher order than the second. We shall, however, limit our discussions to those given.

Just as for the integrals already discussed, so for double integrals we can define an upper and a lower integral. For example, let $f(x, y)$ be defined for the region $R$, and suppose we form the sum

$$\sum_{k=1}^{n} \bar{f}(\xi_k, \eta_k)\Delta_k(x, y),$$

where for each $k$ the value $\bar{f}(\xi_k, \eta_k)$ is the least upper bound of $f(x, y)$ in the rectangle $\Delta_k(x, y)$. If the limit of this sum exists as $n$ becomes infinite in such a manner that the norm $\Delta(x, y)$ approaches zero, we call this limit the **upper double integral** and denote it by the symbol

$$\overline{\int_R} f(x, y)\, d(x, y).$$

In a similar manner, we may define the **lower double integral**

$$\underline{\int_R} f(x, y)\, d(x, y).$$

A necessary and sufficient condition for the existence in $R$ of the double integral is that the upper and the lower integrals have the same value. If the given function is bounded, it is also necessary and sufficient that the two-dimensional set of points where $f(x, y)$ has an $(x, y)$-discontinuity has the measure zero. The proofs of these statements are similar to those for the corresponding theorems for integrals of functions of a single variable and need not be repeated here.

The double integral has been defined for the continuous domain $R$. One can, of course, generalize this definition by replacing the given region $R$ by a bounded set of points $E$ located in the plane and speak of the integral of $f(x, y)$ with respect to this two-dimensional set.

THEOREM. *If $f(x, y)$ is bounded with respect to the two variables together, and the double integral $\int_R f(x, y)\, d(x, y)$ exists, where $R$ is defined by the inequalities*

$$a \leqq x \leqq b, \quad c \leqq y \leqq d, \tag{1}$$

*then each of the iterated integrals*

$$\int_a^b dx \int_c^d f(x, y)\, dy, \int_c^d dy \int_a^b f(x, y) dx$$

*exists and is equal to the double integral.*

We shall first show that

$$\overline{\int_a^b} dx \overline{\int_c^d} f(x, y)\, dy \leqq \overline{\int_R} f(x, y)\, d(x, y),$$

$$\underline{\int_a^b} dx \underline{\int_c^d} f(x, y)\, dy \geqq \underline{\int_R} f(x, y)\, d(x, y).$$

Let the given region $R$ be divided into smaller rectangles $\Delta_k(x, y)$ by drawing lines parallel to the sides of $R$. From the definition of an upper integral, we have for an arbitrarily small positive number $\epsilon$ and a properly chosen value of $\Delta_k(x, y)$

$$\overline{\int_R} f(x, y)\, d\,(x, y) + \epsilon \geqq \Sigma\, \bar{f}(\xi_k, \eta_k) \Delta_k(x, y),$$

where $\bar{f}(\xi_k, \eta_k)$ is the least upper bound of $f(x, y)$ in $\Delta_k(x, y)$. Denote by $\Delta_{xk}$ the length of the interval on a line parallel to the $Y$-axis and lying in the rectangle $\Delta_k(x, y)$. If this line is a boundary of one or more of the rectangles $\Delta_k(x, y)$, always take the point $(\xi_k, \eta_k)$ in the rectangles lying on the same side of the line. We then have

$$\overline{\int_c^d} f(x, y)\, dy \leqq \Sigma'\, \bar{f}(\xi_k, \eta_k) \Delta_{xk},$$

where $\Sigma'$ denotes the summation taken over those rectangles $\Delta_k(x, y)$ which are cut by the line along which the integral is taken. Forming in a similar manner the upper integral of this result with respect to $x$, we have

$$\overline{\int_a^b} dx \overline{\int_c^d} f(x, y)\, dy \leqq \Sigma\, \bar{f}(\xi_k, \eta_k)\, \Delta_k \leqq \overline{\int_R} f(x, y)\, d\,(x, y) + \epsilon.$$

As this relation holds for all values of $\epsilon$, it follows that

$$\overline{\int_a^b} dx \overline{\int_c^d} f(x, y)\, dy \leqq \overline{\int_R} f(x, y)\, d(x, y).$$

By a similar argument, we can show that

$$\underline{\int_a^b} dx \underline{\int_c^d} f(x, y)\, dy \geqq \underline{\int_R} f(x, y)\, d(x, y).$$

By hypothesis, the double integral exists, and hence we have

$$\overline{\int_R} f(x, y)\, d(x, y) = \underline{\int_R} f(x, y)\, d(x, y).$$

It follows that

$$\overline{\int_R} f(x, y)\, d(x, y) \geqq \underline{\int_a^b} dx \underline{\int_c^d} f(x, y)\, dy \geqq \underline{\int_a^b} dx \underline{\int_c^d} f(x, y) dy$$
$$\geqq \underline{\int_R} f(x, y)\, d(x, y).$$

As the two extreme integrals are equal, it follows that we have

$$\overline{\int_a^b} dx \overline{\int_c^d} f(x, y)\, dy = \underline{\int_a^b} dx \underline{\int_c^d} f(x, y)\, dy = \overline{\int_a^b} dx \overline{\int_c^d} f(x, y)\, dy$$
$$= \underline{\int_a^b} dx \underline{\int_c^d} f(xy)\, dy.$$

Consequently, the iterated integral $\int_a^b dx \int_c^d f(x, y)\, dy$ exists and is equal to the double integral.

Likewise, it may be shown that the iterated integral

$$\int_c^d dy \int_a^b f(x, y) dx$$

exists and is equal to the double integral. The theorem is thus established.

In the statement of the foregoing theorem, we regarded the region $R$ as a rectangle defined by the inequalities (1). If $R$ is not a rectangle, the theorem still holds as we can then circumscribe the given region by a rectangle and define the given function $f(x, y)$ as zero at all points exterior to the given region. We then integrate over the rectangle as before.

As we shall see later, the foregoing theorem gives a sufficient although not a necessary condition; that is, there are functions $f(x, y)$ for which the two iterated integrals exist and are equal but the double integral does not exist.

The existence of one of the iterated integrals does not imply the existence of the other nor of the double integral. For example, consider the function $f(x, y)$ defined as follows:

Let $f(x, y) = \frac{1}{2}$ for all rational values of $y$, and suppose $f(x, y) = x$ for all irrational values of $y$. Then for rational values of $y$ we have

$$\int_0^1 f(x, y)dx = \int_0^1 \frac{1}{2}\, dx = \frac{1}{2}.$$

For irrational values of $y$, we have

$$\int_0^1 f(x, y)dx = \int_0^1 x\, dx = \frac{x^2}{2}\Big]_0^1 = \frac{1}{2}.$$

Hence, we have

$$\int_0^1 dy \int_0^1 f(x, y)dx = \int_0^1 \frac{1}{2}\, dy = \frac{y}{2}\Big]_0^1 = \frac{1}{2}.$$

However, in the region $0 \leqq x \leqq 1, 0 \leqq y \leqq 1, f(x, y)$ is a totally discontinuous function of $y$, except for $x = \frac{1}{2}$, since it has the value $\frac{1}{2}$ for all rational values of $y$ and a value different from $\frac{1}{2}$ for all irrational values. As the measure of the set of these points of discontinuity is not zero, it follows that the integral $\int_0^1 f(x, y)\, dy, x \neq \frac{1}{2}$, can not exist. As these points of discontinuity are also $(x, y)$-discontinuities, it follows also that the double integral does not exist in the given region.

As already stated, the two iterated integrals may exist and be equal and at the same time the double integral taken over the region not exist. One might infer that this would be possible since the iterated integrals depend upon the existence and inequality of two sequential limits, while the double integral depends upon a double simultaneous limit. As we know, the former may exist and be equal without implying the existence of the latter limit. That this inference is correct is shown by the following illustration.[1]

Consider the region $R$ defined by the inequalities

$$0 \leqq x \leqq 1, \qquad 0 \leqq y \leqq 1.$$

---

[1] Cf. Thomae, *Zeitschrift für Math.*, Vol. 23 (1878), p. 67. See also Pringsheim *Ber. der München Ak.*, Vol. 29 (1899), pp. 39–62.

Draw the diagonal $OQ$, Fig. 38. For each point $P$ on this diagonal the two coördinates are equal. We shall consider only those points $P$ of the region $R$ for which both $x$ and $y$ have rational values. For example, let $x_0$ be any rational number $0 \leqq \frac{p}{q} \leqq 1$. Divide the segment $Ox_0$ into $q$ equal parts and through the points

$$0, \frac{1}{q}, \frac{2}{q}, \quad . \quad . \quad . \quad , \frac{p}{q} \qquad (1)$$

draw lines parallel to the $Y$-axis. Through the corresponding points on the $Y$-axis draw lines parallel to the $X$-axis intersecting the first system of lines. On each of these lines there are $p$ such intersections. Let $E$ denote the set of all such intersections as $\frac{p}{q}$ takes all rational values between 0 and 1. It is to be noted, however, that the points contributed to $E$ by any choice of $\frac{p}{q}$ all lie to the left of the line $x_0 = \frac{p}{q}$ and below the line $y_0 = \frac{p}{q}$. It is to be understood that by any subsequent choice of $\frac{p}{q}$ none of the intersections of the additional lines drawn parallel to the two axes with those previously drawn shall be points of $E$.

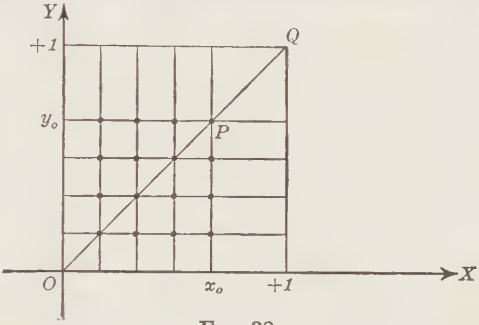


FIG. 38.

The set $E$ has the property that only a finite number of points of $E$ lie on any line in $R$ parallel to either axis. The set $E$ is everywhere dense in $R$; for, it is always possible to select $q$ so that at least one point of $E$ will fall within any subregion of $R$ however small that subregion may be.

We will now define $f(x, y)$ to be zero for all values of $E$ and equal to one for all points of the complementary set $C(E)$. The function $f(x, y)$ is then bounded with respect to the two variables and on any line parallel to either axis it has a finite number of points of discontinuity with respect to $x$ or to $y$, and at each of these points the saltus is one. The integral $\int_0^1 f(x, y)dx$ therefore exists and is equal to one. Hence, we have upon integrating with respect to $y$

$$\int_0^1 dy \int_0^1 f(x, y)dx = 1.$$

In a similar manner, it may be shown that

$$\int_0^1 dx \int_0^1 f(x, y)dy = 1.$$

The double integral $\int_R f(x, y)\, d(x, y)$ can not exist; for, the set $E$ is everywhere dense and at each point the $(x, y)$-saltus is one, and the function $f(x, y)$ is totally discontinuous in the two variables taken together.

It should not be inferred that because the double integral $\int_R f(x, y)\, d(x, y)$ exists the integral $\int_a^b f(x, y)\, dx$ exists for all values of $y$. For, as we know, a double limit may exist without the single limits in the corresponding iterated limit existing for all values of the variables in the neighborhood of the critical point. However, the integral

$$F(y) \equiv \int_a^b f(x, y)dx$$

may not exist for certain values of $y$ and yet both the double integral $\int_R f(x, y)\, d(x, y)$ and the iterated integral $\int_c^d dy \int_a^b f(x, y)\, dx$ may exist, provided the values of $y$ for which $F(y)$ fails to have a definite value form at most a set of measure zero. If it is assumed that $f(x, y)$ is bounded in $(x, y)$, it follows that $F(y)$ is continuous for those values of $y$ for which $f(x, y)$ is continuous in $(x, y)$. From the foregoing theorem, it follows that if the double integral exists, then we have under the above conditions

$$\int_R f(x, y)\, d(x, y) = \int_c^d dy \int_a^b f(x, y)\, dx.$$

## EXERCISES

**1.** Consider the point set discussed in Art. 17 by dropping from the given interval the middle $\frac{1}{3}$, etc. The set $E$ is defined as the end-points of the intervals dropped together with the limiting points of those end-points. For each point of $E$, let $f(x)$ be one, otherwise zero. Show that $f(x)$ is bounded, has points of discontinuity at a set of points having the cardinal number of the continuum and yet the integral $\int_0^1 f(x)dx$ exists.

**2.** Given $f(x) = \underset{k\to\infty}{L}\ \underset{n\to\infty}{L}\ (\cos k!x)^{2n}$. Does the integral $\int_0^1 f(x)dx$ exist? Give reasons for your conclusion.

**3.** Find $\dfrac{dz}{dx}$, where $z = \int_0^1 \log (x^2 + y^2)dy$.

**4.** Find $\dfrac{dz}{dy}$, where $z = \int_{2y}^{y^2} e^{-x} \dfrac{\sin xy}{x}\, dx$.

**5.** Evaluate the following integrals or show that they do not converge.

$$(a)\ \int_0^\infty \frac{x^m dx}{x^n + 1},\ m < n - 1, \qquad (b)\ \int_1^\infty \frac{\log x\, dx}{x\sqrt{x^2 - 1}},$$

$$(c)\ \int_0^{\frac{\pi}{2}} \frac{dx}{\sqrt{\tan x}}, \qquad (d)\ \int_0^\infty \frac{\cos x}{\sqrt{x}}\, dx.$$

**6.** Given $f(x) = \phi_1(x) - \phi_2(x)$, where

$$\phi_1(x) = e^x,\ \phi_2(x) = x^n + a_1x^{n-1} + \cdot\ \cdot\ \cdot + a_n.$$

Show that $f(x)$ is a function of limited variation. What consequences follow from this conclusion?

**7.** Given a function $f(x)$ such that $D^+f(x) < 9$. What properties of $f(x)$ can you state?

**8.** Can a function of limited variation have an infinite number of oscillations in the neighborhood of a point? Could such a function be absolutely continuous? Could it be uniformly continuous? Could it be integrated? Must it be bounded? Give reasons for each conclusion.

**9.** Could a bounded, integrable function have points of discontinuity at a set of points of the first species? first category? at a set having the cardinal number of the continuum? Could such a function belong to Baire's first class? second class?

**10.** Could a function of limited variation be discontinuous at the rational points? at the irrational points? Illustrate your conclusions.

**11.** Test the following integrals for convergence.

$$(a)\ \int_0^\infty e^{-\left(t^2+\frac{1}{t^2}\right)}\, dt, \qquad (b)\ \int_0^\infty y^{\frac{1}{y}}\, dy,$$

$$(c)\ \int_0^2 \frac{\log x}{2 - x}\, dx, \qquad (d)\ \int_0^1 \frac{x^{n-1}}{1 - x}\, dx.$$

**12.** Show that $\cos\left\{\sin\frac{1}{x} + e^{\frac{1}{x}}\right\}$ is an integrable function in the interval $-1 \leqq x < 1$. Is the function $\phi(x) = \left\{\sin\frac{1}{x}\right\}^{e^{\frac{1}{x}}}$ also integrable in the same interval?

**13.** Find the derivative with respect to $y$ of the following functions.

$$\phi_1(y) = \int_{y^2}^{y^3} \frac{x^y - 1}{\log x}\, dx, \quad \phi_2(y) = \int_0^1 x^y \log^n x\, dx.$$

**14.** Given the function $f(x) = \sin\frac{1}{x}$, $x \neq 0$, and let $f(x) = 0$ for $x = 0$. This function has an infinite number of oscillations in the neighborhood of the origin. Is it integrable in the interval $(-1, 1)$? Would a function be integrable if it had the same kind of oscillations in the neighborhood of the points of the set $\left\{\frac{1}{n}\right\}$? at the rational points? at the irrational points? at the points corresponding to the algebraic numbers?

**15.** Without integrating, show that $f(y) = \int_{\frac{1}{2}}^1 y^2 \log x\, dx$ is a continuous function.

**16.** Given $f(x) = \underset{n\to\infty}{L}\left(\cos\frac{\pi}{x}\right)^{2n}$ for $x \neq 0$, and let $f(0) = 0$. Does the integral $\int_{-1}^{+1} f(x)dx$ exist?

**17.** Given $f(x)$ defined by the sequence $\{f_n(x)\}$, where $f_n(x) = \frac{nx}{1 + n^3x^4}\cdot$ Does the integral $\int_{-1}^{+1} f(x)\,dx$ exist? Does the given sequence converge uniformly? quasi-uniformly? Is $f(x)$ continuous?

**18.** The given function $f(x)$ is defined by the sequence $\{f_n(x)\}$, where $f_n(x) = 2n^2x\,e^{-n^2x^2}$. Discuss the existence of the integral $\int_0^1 f(x)dx$. Can we write $\underset{n\to\infty}{L}\int_a^x f_n(x)dx = \int_a^x f(x)dx$?

**19.** Test the following integrals for convergence.

$$(a)\ \int_0^{\infty} e^{-x}\sin\frac{1}{x}\,dx, \qquad (b)\ \int_0^1 \log x \sin\frac{1}{x}\,dx.$$

**20.** Show that every bounded, monotone function can have at most an enumerable set of points of discontinuity.

## CHAPTER VI

## LEBESGUE AND OTHER INTEGRALS

**49. Measurable functions.** As we have seen, the Riemann integral depends for its definition and properties upon dividing the interval or region of integration into a finite number of subdivisions. We shall now consider a more general type of integral based upon the Lebesgue measure of the set of points over which the integral is to be taken. As preliminary to that discussion, we shall introduce what is called a measurable function. For this purpose, let us suppose that we have given a set of points $E$ on the $X$-axis, for which $f(x)$ is single-valued and either finite or infinite with a definite sign. Then $f(x)$ is called a **measurable function** on the set $E$ if $E$ itself is measurable and furthermore if the points of $E$ for which $f(x) \geqq A$ form a measurable set, whatever value is given to $A$. This definition of a measurable function can be extended to a set of points in a space of any number of dimensions. Likewise the theorems concerning measurability may be extended to space of two or more dimensions.

For convenience, we shall denote by

$$E[f(x) < A],\ E[f(x) > A]$$

the subset of $E$ for which $f(x)$ is less than and greater than $A$, respectively. As an illustration of a measurable function, let $f(x)$ be defined for a set $E$ and bounded for the points of this set. The function $f(x)$ is measurable on $E$ if the subset $E[A \leqq f(x) < B]$ is measurable for all values which may be assigned $A$ and $B$. As will be seen, there is no restriction on $f(x)$ as to continuity and no restriction on the set $E$ other than that it must be a measurable set. Among the properties of measurable functions, we have those given by the following theorems.

Theorem I. *The function $f(x)$ is measurable on the measurable set $E$ if any one of the four sets*

$$E[f(x) \geqq A],\ E[f(x) < A],\ E[f(x) > A],\ E[f(x) \leqq A] \qquad (1)$$

*is measurable for every value $A$; moreover, all four sets are then measurable.*

Let us assume that the set $E[f(x) \geqq A]$ is measurable for any value of $A$. It follows that $E[f(x) < A]$ is measurable since it is the difference between the measurable set $E$ and the measurable set $E[f(x) \geqq A]$. For a similar reason, if either of the last two sets is measurable, the other is also.

It remains to show that one of the last two sets is measurable. From the given hypothesis it follows that each of the sets

$$E_n\left[f(x) \geqq A + \frac{1}{n}\right], \quad n = 1, 2, \cdots,$$

is measurable. Since the set $E_n$ is measurable and lies in a finite interval, it follows from Theorem V, Art. 21, that the sum of the sequence of sets $\{E_n\}$ is also measurable. However, the sum of the sequence $\{E_n\}$ is none other than the set $E[f(x) > A]$, and hence this set is measurable. With this result the proof is complete.

We have the following corollary of the foregoing theorem.

COR. 1. *The function $f(x)$ is measurable on the set $E$ if one of the sets* (1), *say $E[f(x) \geqq A]$, is measurable for all rational values of $A$.*

Any real value of $A$, rational or irrational, can be represented as the limit of a sequence of decreasing rational numbers $A_1, A_2, \cdots, A_n, \cdots$. The sequence of sets $\{E_n\}$, where

$$E_n[f(x) \geqq A_n], \quad n = 1, 2, 3 \cdots,$$

satisfies the conditions of Theorem VIII of Art. 21 and consequently the sum of this sequence is measurable, the measure being the limit of the measures of the above sequence of measurable sets. However, the limit of this sequence is the set $E[f(x) > A]$, and this is one of the sets given in (1). Hence, it follows from the foregoing theorem that $f(x)$ is measurable on the set $E$, as the corollary states.

THEOREM II. *If $f(x)$ is measurable on each of a finite or enumerably infinite set of distinct point sets $\{E_n\}$ contained in a finite interval, then $f(x)$ is measurable on the sum $E$ of these sets.*

We have $\quad E = E_1 + E_2 + \cdots + E_n + \cdots.$

By Theorem V, Art. 21, the set $E$ is measurable. Since $f(x)$ is measurable on each $E_n$ it follows that for any $A$, the subset $E_n[f(x) > A]$ is measurable. Then by Theorem VI, Art. 21, we have

$$mE[f(x) > A] = mE_1[f(x) > A] + mE_2[f(x) > A] + \cdots + mE_n[f(x) > A] + \cdots;$$

that is, $f(x)$ is measurable on the set $E$.

THEOREM III. *If $f(x)$ is measurable in a domain $D$ containing the measurable set $E$, then $f(x)$ is measurable on $E$.*

By hypothesis the points of the domain $D$ where $f(x) \geqq A$ form a measurable set $D[f(x) \geqq A]$. Since $E$ is also measurable, the product

$$\{E \cdot D[f(x) \geqq A]\}$$

must be measurable. But this product is nothing else than the set $E[f(x) \geqq A]$, and hence $f(x)$ is measurable on $E$ as the theorem requires.

We have also the following theorem.

THEOREM IV. *If $f(x)$ is measurable on the set $E$, then it is measurable in any finite domain, provided that in this domain it takes the value zero in $C(E)$.*

For, if $f(x)$ is measurable on $E$, then $E$ is measurable and as the given domain $D$ is finite, $C(E)$ is measurable. Since $f(x)$ takes the value zero at every point of $C(E)$, it follows that $C(E)[f(x) \geqq A]$ is measurable. But

$$D[f(x) \geqq A] = E[f(x) \geqq A] + C(E)[f(x) \geqq A]$$

and hence $f(x)$ is measurable in $D$.

THEOREM V. *If $f(x)$ takes a constant value $c$ at all points of a measurable set $E$, then $f(x)$ is measurable on $E$.*

This theorem follows at once from the definition of measurability, since the points of $E$ for which $f(x) \geqq A$ are identical with $E$ for $A \leqq c$ and zero for $A > c$. Hence, in either case the set of points for which $f(x) \geqq A$ is measurable. Consequently, any constant $c$ is a measurable function.

THEOREM VI. *If $f_1(x)$ and $f_2(x)$ are finite and measurable on $E$, then $E[f_1(x) > f_2(x)]$ is measurable.*

For each value of $x$ for which $f_1(x) > f_2(x)$ there is a rational number $r$ lying between $f_1(x)$ and $f_2(x)$. For any rational number $r$ the set of points $x$ for which $f_1(x) > r > f_2(x)$ is the product of the sets $E[f_1(x) > r]$, $E[f_2(x) < r]$ and hence is a measurable set, since it is the product of two such sets. The sum of these products for all values of $r$ is the required set. As the aggregate of rational numbers is an enumerable set, it follows that the required set $E[f_1(x) > f_2(x)]$ is the sum of an enumerable set of measurable sets and is therefore itself measurable.

THEOREM VII. *If $f(x)$ is measurable on a set $E$, then $f(x) \pm c$ and $c\,f(x)$ are also measurable on $E$, where $c$ is any constant.*

We may write

$$E[f(x) + c \geqq A_1] = E[f(x) \geqq A_1 - c].$$

But from the definition of a measurable function, this latter set is measurable, since $E[f(x) \geqq A]$ must be measurable for all values of $A$ and hence when

$$A = A_1 - c.$$

Consequently, $E[f(x) + c \geqq A_1]$ is also measurable. A similar proof establishes the fact that $E[f(x) - c \geqq A_1]$ is measurable.

We may also write

$$E[cf(x) \geqq A_2] = E\left[f(x) \geqq \frac{A_2}{c}\right].$$

If now we set

$$\frac{A_2}{c} = A,$$

it follows from the definition of a measurable function that $E\left[f(x) \geqq \frac{A_2}{c}\right]$ and hence $E[cf(x) \geqq A_2]$ is measurable.

COR. *If $f(x)$ is measurable on the set $E$, then $-f(x)$ is also measurable on that set.*

This proposition follows directly from the foregoing theorem by putting

$$c = -1$$

in the product $c \cdot f(x)$.

We may generalize Theorem VI as follows.

THEOREM VIII. *If $f_1(x)$ and $f_2(x)$ are finite and measurable on a set $E$, then $f_1(x) \pm f_2(x)$ and $f_1(x) \cdot f_2(x)$ are measurable on $E$.*

We have

$$E[f_1(x) \pm f_2(x) > A] = E[f_1(x) > A \mp f_2(x)].$$

The function $A \mp f_2(x)$ is measurable by Theorem VII. Consequently, it follows from Theorem VI that $E[f_1(x) > A \mp f_2(x)]$ and hence $E[f_1(x) \pm f_2(x) > A]$ is measurable.

It remains to show that $f_1(x) \cdot f_2(x)$ is measurable. We know that $\{f(x)\}^2$ is measurable; for, since $A \geqq 0$, we have

$$E[\{f(x)\}^2 \geqq A] = E[f(x) \geqq \sqrt{A}] + E[f(x) \leqq -\sqrt{A}].$$

The two sets in the second member of this equation are measurable, since $f(x)$ is measurable on the set $E$. Consequently, their sum, that is $E[\{f(x)\}^2 \geqq A]$, is also measurable. If now we replace $f(x)$ by $f_1(x) \pm f_2(x)$, it follows that $\{f_1(x) + f_2(x)\}^2$, $\{f_1(x) - f_2(x)\}^2$ are measurable. But we have

$$f_1(x) \cdot f_2(x) = \tfrac{1}{4}[\{f_1(x) + f_2(x)\}^2 - \{f_1(x) - f_2(x)\}^2].$$

It follows from the first part of this theorems and Theorem VII that the second member of this identity is measurable and hence the product in the first member must also be measurable.

It follows from the foregoing theorems that every polynomial is a measurable function.

THEOREM IX. *The upper boundary function $\phi(x)$ of an enumerable set $\{f_n(x)\}$ of measurable functions is measurable. A corresponding statement holds for the lower boundary function.*

We should observe how the function $\phi(x)$ is defined. For each value $x_0$ we take the least upper bound of the set of values $\{f_n(x_0)\}$. The aggregate of these values for all values of $x$ in the interval under consideration is called the upper boundary function $\phi(x)$. It is assumed that for each $x_0$ a finite upper bound of $\{f_n(x_0)\}$ exists; for, otherwise the upper bound $\phi(x)$ would not exist. Suppose now that the functions $f_n(x)$, $n = 1, 2, 3 \cdot \cdot \cdot$ are measurable on the set $E$. Then, we may write

$$E[\phi(x) > A] = E[f_1(x) > A] + E[f_2(x) > A] + \cdot \cdot \cdot \\ + E[f_n(x) > A] + \cdot \cdot \cdot \cdot. \quad (1)$$

Let us put

$$E[\phi_n(x) > A] \equiv E[f_n(x) > A] - \{E[f_1(x) > A] + \cdot \cdot \cdot \\ + E[f_{n-1}(x) > A]\}, n = 1, 2, 3 \cdot \cdot \cdot .$$

The first member of this equation is then measurable, since it is the difference of two measurable sets. Equation (1) is equivalent to the following:

$$E[\phi(x) > A] = E[\phi_1(x) > A] + E[\phi_2(x) > A] + \cdot \cdot \cdot \\ + E[\phi_n(x) > A] + \cdot \cdot \cdot \cdot .$$

The terms of the second member are measurable and satisfy the conditions of Theorem VI, Art. 21, and hence $\phi(x)$ is a measurable function.

COR. *If a monotone sequence $\{f_n(x)\}$ of measurable functions has a limit, then the limiting function is measurable.*

Let us consider first the case in which the given sequence is monotone increasing. For each value of $x$, we then have

$$f_1(x) \leqq f_2(x) \leqq f_3(x) \leqq \cdot \cdot \cdot \leqq f_n(x) \leqq \cdot \cdot \cdot . \qquad (2)$$

The limiting function of this sequence is also the upper boundary function of the set $\{f_n(x)\}$ and hence is measurable by the foregoing theorem. A corresponding proof holds for a monotone decreasing sequence.

THEOREM X. *The upper and the lower limiting functions of any sequence of measurable functions are measurable.*

We shall prove the theorem for the upper limiting function. A similar proof holds for the lower limiting function. Denote the given sequence by $\{f_n(x)\}$. For any value $x_0$ let $F(x_0)$ be the upper limit of the sequence $\{f_n(x_0)\}$. Then $F(x)$ will denote the upper limiting function of the given sequence $\{f_n(x)\}$. Denote by $\phi_n(x)$ the upper boundary function of the sequence

$$f_n(x), f_{n+1}(x), \cdot \cdot \cdot, \quad n = 1, 2, 3, \cdot \cdot \cdot .$$

Then the sequence $\{\phi_n(x)\}$ will have a lower boundary function which is the upper limiting function of the sequence $\{f_n(x)\}$. Each of the functions $\phi_n(x)$ is measurable since it is the upper boundary function of an enumerable set of measurable functions. Likewise the lower boundary function of the set $\{\phi_n(x)\}$ of measurable functions is measurable. But, as we have seen, this lower boundary function is the upper limiting function of the given set $\{f_n(x)\}$, hence the theorem.

COR. *A limiting function of a sequence of measurable functions is measurable.*

This corollary follows from the foregoing theorem; for, when the given functions have a unique limiting function that limiting function is the common value of the upper and the lower limiting functions. Moreover, since the limiting function exists, it is the limiting function of any enumerable sequence of the given functions.

Since every polynomial is measurable, it follows from this corollary that every continuous function is measurable, since every such function can be represented as the limit of a sequence of polynomials. It then follows that every function of Baire's class one is measurable, since every such function can be represented as the limiting function of a sequence of continuous functions; that is, every point-wise discontinuous function is measurable. It may also be shown that a function belonging to any one of Baire's classes is measurable.[1]

In the Lebesgue theory of integration, it is of great importance that one is able to say that the upper and lower limiting functions of a sequence $\{f_n(x)\}$ of measurable functions are measurable even though the limiting function has points of discontinuity which may be finite or infinite. The following is an illustration. Let

$$f_n(x) = \frac{n + n^2x^2}{1 + n^2x^2}.$$

Then we have for the limiting function, since the upper and the lower limiting functions are equal,

$$\begin{aligned} f(x) &= \underset{n\to\infty}{L} \frac{n + n^2x^2}{1 + n^2x^2} = 1, \text{ for } x \neq 0, \\ &= \infty, \text{ for } x = 0. \end{aligned}$$

The limiting function $f(x)$ is, however, measurable by the foregoing corollary, since it is a limiting function of a sequence of measurable functions. That $f(x)$ is measurable follows also from the definition of a measurable function.

THEOREM XI. *The function $f(x)$ is measurable on the interval $(a, b)$, provided the points of discontinuity of $f(x)$ form at most a set of measure zero.*

Consider the set of points $E$ of $(a, b)$ where $f(x) \geqq A$, $A$ being any finite number. All the limiting points of $E$ which are not at the same time points of $E$ must be points of discontinuity of $f(x)$. By adding these points to $E$ we have a bounded, closed set $E'$, which by Theorem III, Art. 21 must be a measurable set. By hypothesis the set of points of discontinuity $G$ of $f(x)$ has the measure zero. Hence, the set $E_1$, where

$$E_1 = E' - E,$$

is also of measure zero, as it is a subset of $G$. The set $E$ is measurable, since it is the difference of two measurable sets, and consequently the function $f(x)$ is measurable on $(a, b)$.

---

[1] See Lebesgue, *Leçons sur l'intégration*, p. 112.

THEOREM XII. *Suppose $f(x, y)$ to be a single-valued function of $x$ and $y$ in the region $a \leqq x \leqq b$, $y_0 < y \leqq d$. Moreover, let $f(x, y)$ be continuous in $y$ and measurable in $x$ for each $y \neq y_0$. Then the upper and the lower limiting functions of $f(x, y)$ as $y$ approaches $y_0$ are measurable functions in $x$*

For convenience, put

$$\phi_1(x) \equiv \overline{\underset{y \to y_0}{L}} f(x, y), \quad \phi_2(x) \equiv \underline{\underset{y \to y_0}{L}} f(x, y).$$

In this discussion it is assumed that $\phi_1(x)$, $\phi_2(x)$ exist and are finite. We shall show that $\phi_1(x)$ is a measurable function. A similar proof may be readily formulated for the lower limiting function $\phi_2(x)$.

For any given value of $x$, say $x_0$, the function $f(x_0, y)$ has a least upper bound in the interval $(y_0, y_0 + \epsilon)$. As $f(x, y)$ is continuous in $y$ this least upper bound is also the upper limit and may be obtained by considering only the rational values of $y$ in the interval. The limit, as $\epsilon$ approaches zero, of this least upper bound is the value of $\phi_1(x)$ for $x = x_0$, and the aggregate of these values for all values of $x$ in $(a, b)$ constitutes the upper limiting function $\phi_1(x)$. Let $\{y_n\}$ be the set of rational values of $y$ in the given interval. By hypothesis each function of the sequence $\{f(x, y_n)\}$ is measurable in $x$ on the interval $(a, b)$. From the corollary to Theorem X, it follows that the limiting function $\phi_1(x)$ is likewise measurable.

As an important consequence of the foregoing theorem, we have the following corollaries.

COR. 1. *Every derivative number of a continuous function is measurable.*

We shall establish this corollary for the upper right-hand derivative number. A similar proof holds for the other derivative numbers. For convenience, we shall put

$$F(x, \Delta x) \equiv \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad a \leqq x \leqq b, \quad \Delta x > 0.$$

The derivative number $D^+ f(x)$ is then the limit of the least upper bound of $F(x, \Delta x)$ as $\Delta x$ approaches zero through positive values. For $\Delta x \neq 0$, $F(x, \Delta x)$ is continuous in $x$ and in $\Delta x$. Moreover, it is measurable in $x$ for each value of $\Delta x > 0$ by virtue of Theorem IX. The function $F(x, \Delta x)$ satisfies the conditions of the foregoing theorem, and hence its upper limit as $\Delta x$ approaches zero, that is $D^+ f(x)$, is measurable.

As an illustration, let

$$\begin{aligned} f(x) &= x^2 \sin \frac{1}{x^2}, \quad x \neq 0 \\ &= 0 \qquad\qquad\;\, , \quad x = 0. \end{aligned}$$

This function is continuous and its derivative function is

$$\begin{aligned} f'(x) &= 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2}, \quad x \neq 0, \\ &= 0, \quad x = 0, \end{aligned}$$

which by the foregoing corollary must be a measurable function, although it is not bounded.

COR. 2. *Every finite derivative function is measurable.*

If $f'(x)$ is such a derivative function, then $f(x)$ is continuous and by Corollary 1 every derivative number of $f(x)$ is measurable. Consequently, their common values, namely the derivative function, must also be measurable.

The foregoing theorem gives the measure of a limiting function $f(x, y_0)$ considered as a function of $x$. The notion of measurable functions may be readily extended to functions of two or more variables. Thus, $f(x, y)$ is said to be measurable on the two-dimensional set $E$ if $E$ itself is measurable and furthermore if the points of $E$ for which

$$f(x, y) \geqq A$$

form a measurable set, $A$ being any finite number however large. The general properties of measurable functions of a single variable which are developed in this article can be extended to functions of two variables. Thus the sum or the product of two such measurable functions is measurable; any function belonging to one of Baire's classes is a measurable function, etc.

**50. Definition of a Lebesgue integral.** As already pointed out, the Lebesgue integral is based upon the properties of measurable functions and the measure of a point set. Let $f(x)$ be a single-valued, bounded, measurable function defined for the interval $(a, b)$. For the present we shall assume that $f(x)$ is also positive for all values of $x$ in the given interval, and later we shall show how the definition can be extended to negative functions and to the more general case where the integral is taken over a set of points. Let $y = c$ and $y = d$ be the lower and the upper bounds, respectively, of $f(x)$ in the given interval. Instead of dividing the interval $(a, b)$ on the $X$-axis into $n$ subintervals, as was done in the case of the Riemann integral, let us divide the interval $(c, d)$ on the $Y$-axis into $n$ subintervals by the insertion of the intermediate points $y_1, y_2, \cdots, y_{k-1}, y_k, \cdots y_{n-1}$, as indicated in Fig. 39.

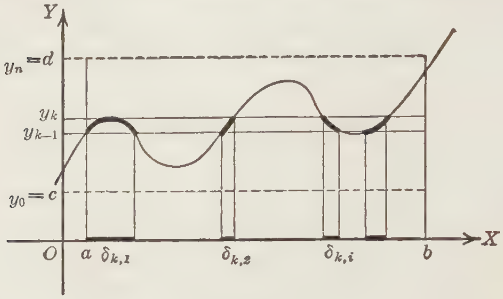


Fig. 39.

Let $E_k$ be the set of points in the interval $(a, b)$ for which

$$y_{k-1} \leqq f(x) < y_k.$$

As we shall see, the Lebesgue integral may be considered as the measure of a two-dimensional point set. Denote by $\eta_k$ any value of $y$ in the interval $(y_{k-1}, y_k)$. Let

$$\epsilon > (y_k - y_{k-1}),\ k = 1, 2, \cdots, n. \tag{1}$$

We may now define the Lebesgue integral analytically for the positive interval $(a, b)$ by the relation

$$\int_a^b f(x)dx = \underset{\epsilon \to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_k), \tag{2}$$

where, in this case, $m(E_k) = \sum\limits_{i=1}^{p} \delta_{ki}$,

as seen from the figure. It will be observed that as $\epsilon$ approaches zero, the value of $n$ increases indefinitely although the converse may not be true.

We now complete the definition for bounded, measurable functions by setting up the convention that

$$\int_a^b f(x)dx = -\int_b^a f(x)dx.$$

Instead of defining $f(x)$ for the interval $(a, b)$, it might be defined for any measurable set $E$ of points on the $X$-axis. In the foregoing definition $m(E_k)$ is the measure of the subset of $E$ for which

$$y_{k-1} \leqq f(x) < y_k.$$

If the function $f(x)$ satisfies this definition, it is said to be integrable in the Lebesgue sense over the set $E$, where

$$E \equiv \underset{\epsilon \to 0}{L} \sum_{k=1}^{n} E_k.$$

We shall indicate this integral by the symbol

$$\int_E f(x)dx.$$

As we shall see, the effectiveness of the Lebesgue integral as compared with the Riemann integral is much increased by the fact that it may be associated in its applications with functions which are defined for a set of points instead of an interval.

The restriction that $f(x)$ shall have only positive values can now be removed; for, all we need to do is to consider separately the set of points $\{x\}$ for which $f(x)$ is negative and then take the algebraic sum of this result and the one for positive values of $f(x)$.

We can now establish the following existence theorem.

THEOREM. *The Lebesgue integral* $\int_E f(x)dx$ *exists if* $f(x)$ *is measurable and bounded on the point set* $E$.

To establish this theorem, we must show that the limit given in (2) exists under the conditions set forth in the theorem and that the limiting value is independent of the manner in which the point set $\{y_k\}$ is chosen.

Let
$$\phi(n) \equiv \sum_{k=1}^{n} y_{k-1}\, m(E_k), \quad \psi(n) \equiv \sum_{k=1}^{n} y_k\, m(E_k). \tag{3}$$

We then have
$$\phi(n) \leqq \sum_{k=1}^{n} \eta_k m(E_k) \leqq \psi(n). \tag{4}$$

The functions $\phi(n)$, $\psi(n)$ are both bounded and monotone, the first being monotone increasing and the second monotone decreasing as the length of the intervals $(y_{k-1}, y_k)$ decreases by the insertion of intermediate points. Consequently, each function has a limit as $n$ increases indefinitely in such a manner that $\epsilon > y_k - y_{k-1}$ approaches zero. These two limits are equal; for, we have

$$\psi(n) - \phi(n) = \Sigma(y_k - y_{k-1})\cdot m\,(E_k) < \epsilon\cdot m(E).$$

Since $\epsilon$ is arbitrarily small and $E$ is measurable and hence $m(E)$ is finite, it follows that $\epsilon\cdot m(E)$ is arbitrarily small. Consequently, we have
$$\underset{\epsilon\to 0}{L}\, \{\psi(n) - \phi(n)\} = 0. \tag{5}$$

By Theorem IV, Art. 24, it follows from (4) that the limit

$$\underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_k)$$

exists and is equal to the common limit

$$\underset{\epsilon\to 0}{L}\, \phi(n) = A = \underset{n\to\infty}{L}\, \psi(n). \tag{6}$$

It remains to show that the value of this limit is independent of the choice of the set of values $\{y_k\}$. Let $\{y_k'\}$ be any other set of points in the interval $(c, d)$. Superimpose the set $\{y_k'\}$ upon the set $\{y_k\}$. Some of the points of the two sets may coincide. The remaining points of $\{y_k'\}$ may be regarded as the points inserted in the first part of this discussion in establishing the limits (6). Hence, if we form the functions $\phi'(n)$ and $\psi'(n)$ with reference to the set $\{y_n'\}$ as $\phi(n)$ and $\psi(n)$ were formed from $\{y_k\}$, it follows that $\phi'(n)$, $\psi'(n)$ will differ at most by $\epsilon' \, m(E)$, where $\epsilon'$ is arbitrarily small. Denote by $\psi''(n)$, $\phi''(n)$ the corresponding sums formed from the set $\{y_k''\}$ obtained by superimposing $\{y_k\}$ upon $\{y_k'\}$. Then $\psi''(n)$, $\phi''(n)$ lie between $\psi(n)$ and $\phi(n)$ and likewise between $\psi'(n)$ and $\phi'(n)$. Consequently, $\psi''(n)$ can differ from $\psi(n)$ at most by $\epsilon \; m(E)$ and from $\psi'(n)$ at most by $\epsilon' \, m(E)$. It follows that $\psi(n)$ and $\psi'(n)$ can differ at most by $(\epsilon + \epsilon') \, m(E)$. The same is true of $\phi(n)$ and $\phi'(n)$. Hence, as the functions approach coincidence, the functions $\phi'(n)$, $\psi'(n)$ must approach the same limiting value $A$. The value of the limit is therefore not dependent upon the particular manner in which the set of points $\{y_k\}$ was chosen. With this conclusion the theorem is established.

In case one knows that the integral exists, its value may be determined by either of the limits in (6); that is, by either $\underset{n\to\infty}{L} \phi(n)$ or $\underset{n\to\infty}{L} \psi(n)$.

If $f(x)$ is defined for all values of $x$ in a given interval $(a, b)$, is continuous throughout this interval, and has only a finite number of maxima and minima, it follows from the foregoing theorem that the Lebesgue integral exists. As we have already seen, Art. 39, the Riemann integral also exists. That the two are equal follows from the definition given in (2). For, $\eta_k$ can be replaced by the functional value $f(\xi_k)$, where $y_{k-1} \leqq f(\xi_k) \leqq y_k$, and in this case the measure $m(E_k)$ is the sum of the $x$-intervals composing $E_k$. We have then

$$\underset{\Delta x\to 0}{L} \sum_{k=1}^{n} f(\xi_k) \, \Delta_k x = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k \, m(E_k).$$

Under the conditions imposed upon $f(x)$, both of these limits exist, the first defining the Riemann integral taken over the interval $(a, b)$ while the second defines the Lebesgue integral over the same interval.

The equality of the two integrals is thus shown for a very special class of functions. Later we shall see that any bounded function which is integrable in the Riemann sense is also integrable in the

Lebesgue sense and the two integrals are the same. The converse of this proposition, it will be seen, is, however, not true.

It was pointed out in the last article that every function belonging to one of Baire's classes is a measurable function. We can now say, as a consequence of the foregoing theorem, that every such function is integrable in the Lebesgue sense, provided it is also bounded.

One can readily formulate the corresponding definition of a Lebesgue integral of two or more variables. Suppose, for example, we have given a bounded, measurable function $f(x, y)$ of two independent variables defined for a two-dimensional point set $E$. Then one may define the Lebesgue double integral with respect to $E$ by the relation

$$\int\int_E f(x, y)dxdy = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_k),$$

where in this case $E_k$ is a two-dimensional subset of $E$ for which $z = f(x, y)$ satisfies the conditions

$$z_{k-1} \leqq f(x, y) < z_k, \; | z_k - z_{k-1} | < \epsilon,$$

and $\eta_k$ is any value measured on the $Z$-axis and situated in the interval $(z_k - z_{k-1})$. The existence proof stated in the foregoing theorem can be extended to multiple-valued integrals. The proof is analogous to that given for the foregoing theorem and is left as an exercise for the reader.

**51. Lebesgue integrals of bounded functions.** In the present article we shall consider only bounded functions. As we shall see, the Lebesgue integrals have certain properties analogous to those of the Riemann integrals. Among these we have those given in the following theorems:

THEOREM I. *If the Lebesgue integral* $\int_E f(x)dx$ *exists, then we have* $\int_E c\cdot f(x)dx = c\int_E f(x)dx$, *where* $c$ *is any constant.*

In the relation given by (2) Art. 50, the factor $\eta_k$ can be replaced by $c\eta_k$ and hence by the laws of limits, we have

$$\underset{\epsilon\to 0}{L} \sum_{k=1}^{n} c\, \eta_k\, m(E_k) = c \underset{\epsilon\to 0}{L} \sum \eta_k\, m(E_k);$$

that is,

$$\int_E cf(x)dx = c\int_E f(x)dx, \tag{1}$$

which is the desired result.

The Law of the Mean for Lebesgue integrals may be stated in the following theorem.

THEOREM II. *If* $f(x)$ *is bounded and measurable on the set* $E$ *having respectively* $A$ *and* $B$ *as the least upper and greatest lower bounds, then*

$$B\cdot m(E) \leqq \int_E f(x)dx \leqq A\cdot m(E). \tag{2}$$

From the discussion in the previous article, we have

$$B \leqq \eta_k,\ A \geqq \eta_k,\ m(E) = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} m(E_k).$$

It then follows that

$$B\cdot m(E) \leqq \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k m(E_k) \leqq A\cdot m(E),$$

whence from the definition of a Lebesgue integral the required relation (2) follows.

In case $A = B = 1$, we have the special relation

$$\int_E dx = m(E). \tag{3}$$

THEOREM III. *If* $f(x)$ *is bounded and measurable on a finite number or an infinite sequence of distinct, measurable point sets* $\{E_n\}$ *whose sum is* $E$, *then*

$$\int_E f(x)dx = \int_{E_1} f(x)dx + \int_{E_2} f(x)dx + \cdot\ \cdot\ \cdot + \int_{E_n} f(x)dx + \cdot\ \cdot\ \cdot\ . \tag{4}$$

Let us first consider the sum of two integrals, that is where $E$ is the sum of the two point sets $E_1$, $E_2$ having no points in common. From the definition of a Lebesgue integral we have

$$\int_{E_1} f(x)dx = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_{1k}),$$

$$\int_{E_2} f(x)dx = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_{2k}).$$

By addition, we obtain

$$\int_{E_1} f(x)dx + \int_{E_2} f(x)dx = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k[m(E_{1k}) + m(E_{2k})]$$

$$= \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k\, m(E_k), \text{ by Theorem VI, Art. 21,}$$

$$= \int_E f(x)dx. \tag{5}$$

Instead of the sum of two point sets, we may by the same method extend the foregoing result to the sum of any finite number of distinct sets.

We shall now consider the case in which we have the sum of an infinite sequence of distinct, measurable sets of points. Denote by $S_n$ the sum of the first $n$ sets and by $R_n$ the sum of the remaining sets. The sets $S_n$, $R_n$ are both measurable by Theorem V of Art. 21. Moreover, since $f(x)$ is measurable on each $E_n$ it is by Theorem II, Art. 49, measurable on $S_n$ and $R_n$. Then by the first part of the above theorem, we have

$$\int_E f(x)dx = \int_{S_n} f(x)dx + \int_{R_n} f(x)dx. \tag{6}$$

By Theorem X, Art. 21, it follows that

$$\underset{n\to\infty}{L}\, m(R_n) = m(\underset{n\to\infty}{L}\, R_n) = 0.$$

By aid of Theorem II, we therefore have

$$\underset{n\to\infty}{L} \int_{R_n} f(x)dx = 0.$$

Then from (6) it follows that

$$\begin{aligned}\int_E f(x)dx &= \underset{n\to\infty}{L} \int_{S_n} f(x)dx \\ &= \int_{E_1} f(x)dx + \int_{E_2} f(x)dx + \\ &\quad \cdots + \int_{E_n} f(x)dx + \cdots,\end{aligned}$$

which was to be proved.

It is to be noted that the foregoing theorem gives a generalization of a well-known theorem concerning Riemann integrals, namely:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx,$$

where $a < c < b$.

It is also of interest to note that, while the Riemann integral is additive over a finite number of domains of integration, the Lebesgue integral is additive over an enumerably infinite number of subdivisions of the given domain.

Theorem IV. *If $f_1(x)$ and $f_2(x)$ are bounded and measurable on the set $E$ and satisfy the relation $|f_1(x)| \leqq |f_2(x)|$, then*

$$\int_E |f_1(x)|\, dx \leqq \int_E |f_2(x)|\, dx.$$

For a given $\eta_k$, let $E$ be decomposed into $n$ distinct, measurable subsets $E_1, E_2, \ldots E_k, \ldots, E_n$, such that the greatest lower bound of $|f_1(x)|$ on $E_k$ is $\eta_k$. Since $f_2(x)$ is bounded and measurable on each $E_k$ as well as on $E$, it follows by Theorem III that

$$\int_E |f_2(x)|\,dx = \sum_{k=1}^{n} \int_{E_k} |f_2(x)|\,dx. \tag{7}$$

However, from the conditions stated in the theorem, we have

$$\int_{E_k} |f_2(x)|\,dx \geqq \eta_k m(E_k). \tag{8}$$

Since the integral $\int_E f_1(x)dx$ exists, it follows from the definition of a Lebesgue integral that

$$\int_E |f_1(x)|\,dx = \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k m(E_k). \tag{9}$$

By taking the sum from $k = 1$ to $k = n$ of (8) and then allowing $\epsilon$ to approach zero as $n$ becomes infinite, we have from (7), (8) and (9)

$$\int_E |f_2(x)|\,dx \geqq \underset{\epsilon\to 0}{L} \sum_{k=1}^{n} \eta_k m(E_k) = \int_E |f_1(x)|\,dx,$$

as stated in the theorem.

THEOREM V. *If $f(x)$ is bounded and measurable on $E$, then*

$$\left|\int_E f(x)dx\right| \leqq \int_E |f(x)|\,dx.$$

Denote by $E_1$ that subset of $E$ where $f(x)$ is positive or zero and by $E_2$ that subset where $f(x)$ is negative. Then the integrals $\int_{E_1} f(x)dx$, $\int_{E_2} f(x)dx$ are either zero or of opposite signs. The integral in the second member of the given inequality is the sum of the numerical values of these two integrals while that in the first member is the numerical value of their difference. From this fact the theorem follows.

THEOREM VI. *If each of a finite number of functions $f_k(x)$ is bounded and measurable on a given set $E$, then*

$$\int_E [f_1(x) + \cdots + f_k(x) + \cdots + f_n(x)]dx = \int_E f_1(x)dx + \cdots + \int_E f_k(x)dx + \cdots + \int_E f_n(x)dx.$$

It is sufficient to prove the theorem for the sum of two functions; for, by repeating the process we can extend the results to the sum of any finite number of functions. Each of the given integrals exists; for, since $f_k(x)$ is bounded and measurable on $E$, then $\sum\limits_{k=1}^{n} f_k(x)$ is bounded and by Theorem VIII of Art. 49 it is also measurable.

Let us first consider the integral $\int_E [f(x) + A]dx$, where $A$ is a constant. Denote by $\phi_1(n)$ the particular value of $\phi(n)$ defined by (3) of Art. 50 for $[f(x) + A]$ and by $\phi_2(n)$ its value for $f(x)$. We then have

$$\phi_1(n) = \sum_{k=1}^{n} (y_{k-1} + A)m(E_k) = \sum_{k=1}^{n} y_{k-1}m(E_k) + \sum_{k=1}^{n} Am\,(E_k)$$
$$= \sum_{k=1}^{n} y_{k-1}m(E_k) + Am(E) = \phi_2(n) + Am(E).$$

Upon passing to the limit as in Art. 50, we have

$$\int_E [f(x) + A]dx = \int_E f(x)dx + A \cdot m(E). \tag{10}$$

Let us now consider the integral of the sum of two functions, $f_1(x)$ and $f_2(x)$, each being bounded and measurable on the point set $E$. We have, if we think of $E$ as decomposed into $n$ distinct, measurable subsets $E_k$ corresponding to the functional values $y_{k-1} < f_1(x) \leqq y_k$,

$$\int_E [f_1(x) + f_2(x)]dx \geqq \sum_{k=1}^{n} \int_{E_k} [y_{k-1} + f_2(x)]dx, \tag{11}$$

$$\int_E [f_1(x) + f_2(x)]dx \leqq \sum_{k=1}^{n} \int_{E_k} [y_k + f_2(x)]dx. \tag{12}$$

But by (10) we have

$$\int_{E_k} [y_{k-1} + f_2(x)dx = \int_{E_k} f_2(x)dx + y_{k-1} \cdot m(E_k),$$
$$\int_{E_k} [y_k + f_2(x)]dx = \int_{E_k} f_2(x)dx + y_k \cdot m(E_k).$$

Making use of Theorem III, we have from (11), (12)

$$\int_E f_2(x)dx + \sum_{k=1}^{n} y_k m(E_k) \geqq \int_E [f_1(x) + f_2(x)]dx$$
$$\geqq \int_E f_2(x)dx + \sum_{k=1}^{n} y_{k-1}m(E_k). \tag{13}$$

As $n$ becomes infinite in such a manner that $\epsilon > (y_k - y_{k-1})$ approaches zero, it follows that

$$\sum_{k=1}^{n} y_k m(E_k), \ \sum_{k=1}^{n} y_{k-1} m(E_k)$$

both approach the same limit, namely $\int_E f_1(x)dx$. Hence, from Theorem IV, Art. 24, we have from (13)

$$\int_E [f_1(x) + f_2(x)]dx = \int_E f_1(x)dx + \int_E f_2(x)dx.$$

The theorem is thus established.

The following theorem gives the conditions under which the foregoing theorem may be extended to an infinite sequence of functions.

THEOREM VII. *Let $\{f_n(x)\}$ be an infinite sequence of measurable functions defined for the set of points $E$. Moreover, suppose these functions to be bounded on this set; that is, simultaneously for all values of $x$ in $E$ and all integral values of $n$. If $\{f_n(x)\}$ converges to the unique limiting function $f(x)$ as $n$ becomes infinite, then we have*

$$\underset{n \to \infty}{L} \int_E f_n(x)dx = \int_E f(x)dx.$$

As each function of the sequence $\{f_n(x)\}$ is measurable, it follows from the corollary to Theorem X of Art. 49 that the limiting function $f(x)$ is measurable. It is also bounded, since $f_n(x)$ is bounded when considered as a function of $x$ and $n$ together. In order to establish this theorem, it is sufficient to prove that the integral

$$\int_E [f_n(x) - f(x)]dx$$

converges to zero as $n$ becomes infinite.

As both $f_n(x)$ and $f(x)$ are bounded and measurable, then $\{f_n(x) - f(x)\}$ is also bounded and measurable on the set $E$. It follows that the integral $\int_E [f_n(x) - f(x)]dx$ exists. Denote by $E_n$, $n = 1, 2, 3, \cdots$, the points of $E$ where

$$|f_n(x) - f(x)| < \epsilon. \tag{14}$$

Since $\{f_n(x) - f(x)\}$ is bounded on $E$, we have for the complementary set of $E_n$, namely $C(E_n)$,

$$|f_n(x) - f(x)| < M, \tag{15}$$

where $M$ is a definite number. We may write

$$\left| \int_E [f_n(x) - f(x)]dx \right| \leqq \int_E | f_n(x) - f(x) | \, dx$$
$$\leqq \epsilon m(E_n) + M \cdot mC(E_n). \qquad (16)$$

Upon passing to the limit, we have from Theorem X, Art. 21,

$$\underset{n \to \infty}{L} \; m(E_n) = m(E),$$

and hence

$$\underset{n \to \infty}{L} \; mC(E_n) = 0.$$

Consequently, from (16) we obtain, for arbitrarily small values of $\epsilon$, $\epsilon_1$ and for $n$ sufficiently large,

$$\left| \int_E [f_n(x) - f(x)]dx \right| < \epsilon \cdot m(E) + \epsilon_1,$$

and hence

$$\underset{n \to \infty}{L} \int_E [f_n(x) - f(x)]dx = 0,$$

which establishes the theorem.

The foregoing theorem can be readily extended to functions of two or more variables. This theorem shows an important distinction between the Riemann and the Lebesgue integrals, namely: In the foregoing theorem merely the convergence of the sequence $\{f_n(x)\}$ of bounded functions is sufficient for the conclusion that the additive property of the integral with respect to the integrand shall be valid, while in the case of the Riemann integral, on the other hand, the sequence $\{f_n(x)\}$ must converge in a special manner, namely quasi-uniformly in general.

As a consequence of the foregoing theorem, we may state the following corollary.

COR. *If the sequence $\{f_n(x)\}$ of functions measurable on $E$ has a limiting function $f(x)$, then*

$$\underset{n \to \infty}{L} \int_E f_n(x)dx = \int_E f(x)dx,$$

*provided all the elements $f_n(x)$ of the sequence have the same finite bounds.*

**52. Comparison of Riemann and Lebesgue integrals.** Suppose $f(x)$ to be defined for the interval $(a, b)$, and in this interval let it be single-valued and bounded. The question presents itself as to whether the existence of the Riemann integral implies the existence of the Lebesgue integral and conversely. As already pointed out, an essential difference between the two is that in the Riemann integral the interval of integration is subdivided while in the Lebesgue integral the subdivisions are taken on the axis of the dependent variable in stating the definition of an integral. The answer to the foregoing question for bounded functions is to be found in the following theorem. The case of non-bounded functions will be discussed in the following article.

THEOREM: *Let $f(x)$ be defined for the interval $(a, b)$ and for this interval suppose it to be single-valued, bounded, and integrable in the Riemann sense. It is then integrable in $(a, b)$ in the Lebesgue sense, and the two integrals are equal. The converse, however, is not necessarily true.*

Within the interval $(a, b)$ consider the set of points $E$ where $f(x) > A$, $A$ being any finite value. This set may contain both points at which $f(x)$ is continuous and also points of discontinuity of $f(x)$. From the nature of continuity, each point of $E$ at which $f(x)$ is continuous is an inner point of $E$, since there must exist a neighborhood of such a point in which $f(x) > A$ for all values of $x$. The points of $E$ which are points of continuity form then an open set. Since $f(x)$ is integrable in the Riemann sense, it follows from Art. 42 that the points of $E$ where $f(x)$ is discontinuous form at most a set of measure zero. Then from Theorem V, Art. 21, we know that $E$ is a measurable set; for it is the sum of two measurable sets. Consequently, $f(x)$ is a measurable function on $(a, b)$, and since $f(x)$ is also bounded, it follows by the theorem of Art. 50 that it is integrable in the Lebesgue sense.

We shall now show that the two integrals are equal under the conditions set forth in the theorem. Let the given interval $(a, b)$ be divided into subintervals

$$\Delta_k x \equiv x_k - x_{k-1},$$

and denote by $A_k$, $B_k$ the least upper and the greatest lower bounds, respectively, of $f(x)$ in $\Delta_k(x)$. By Theorem II of the last article the Lebesgue integral satisfies the following condition, since the measure of an interval is its length,

$$B_k \, \Delta_k x \leqq \text{Lebesgue} \int_{x_{k-1}}^{x_k} f(x)dx \leqq A_k \Delta_k x.$$

Summing up the products formed for all the intervals $\Delta_k x$ of $(a, b)$, we have

$$\sum_{k=1}^{n} B_k \, \Delta_k x \leqq \text{Lebesgue} \int_a^b f(x)dx \leqq \sum_{k=1}^{n} A_k \, \Delta_k x.$$

As $n$ becomes infinite in such a manner that the norm $\Delta x$ approaches zero, $\Sigma B_k \Delta_k x$, $\Sigma A_k \Delta_k x$ have as their common limiting value the Riemann integral from $a$ to $b$. From Theorem IV, Art. 24, it then follows that the two integrals are equal for bounded functions whenever the Riemann integral exists.

The converse of the foregoing theorem is not true; that is a Lebesgue integral of a bounded function may exist in an interval $(a, b)$, while the Riemann integral does not exist. For example, let $f(x)$ be defined for the interval $(0, 1)$ as follows:

$$\begin{aligned} f(x) &= 1, \text{ for rational values of } x, \\ &= 0, \text{ for irrational values of } x. \end{aligned}$$

This function is totally discontinuous in the given interval and hence has no integral in the Riemann sense; for, as we know, the necessary and sufficient condition for the existence of such an integral is that the points of discontinuity shall form at most a set of measure zero. In this case the measure of the set of points of discontinuity is one. The function $f(x)$ is bounded and measurable on the set $E_1$ of rational points and also on the set $E_2$ of irrational points. Consequently, by the Theorem of Art. 50, the Lebesgue integrals $\int_{E_1} f(x)dx$, $\int_{E_2} f(x)dx$ both exist. From Theorem III, Art. 51, it follows that the Lebesgue integral $\int_0^1 f(x)dx$ taken over the interval $(0, 1)$ must exist.

**53. Lebesgue integrals for non-bounded functions.** The Lebesgue theory of integration may be extended to certain classes of measurable functions which are not bounded.[1] To do this let $f(x)$ be any measurable, non-bounded function defined on the bounded, measurable set of points $E$. For the present, we shall also assume that $f(x)$ is not negative. Let $k$ be any one of the sequence of positive real numbers,

$$k_1, k_2, k_3, \cdot \cdot \cdot \tag{1}$$

having no finite upper bound. We shall define the auxiliary function $f_k(x)$ by the inequalities

$$\begin{aligned} f_k(x) &= f(x), \text{ where } f(x) \leqq k. \\ &= k, \text{ where } f(x) > k. \end{aligned}$$

The function $f_k(x)$ is then bounded and measurable on the set of points $E$. Consequently, the Lebesgue integral $\int_E f_k(x)dx$ exists for all values of $k$. If the limit of $\int_E f_k(x)dx$ exists as $k$ becomes infinite through the succession of values given in (1), then we say that the integral $\int_E f(x)dx$ exists, and we write

---

[1] See, *Leçons sur l'intégration*, p. 335; also de la Vallée Poussin's *Intégrale de Lebesgue*, p. 45.

$$\int_E f(x)dx = \underset{k\to\infty}{L} \int_E f_k(x)dx. \tag{2}$$

If $f(x)$ is negative for all points of $E$, we say that the Lebesgue integral exists on $E$ if the limit $\underset{k\to\infty}{L} \int_E | f_k(x) | dx$ exists as $k$ becomes infinite through the sequence of values given in (1).

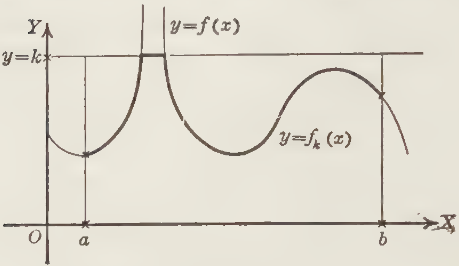


FIG. 40.

If, however, $f(x)$ is negative for some points of $E$ and positive at others, we may define the Lebesgue integral by the relation

$$\int_E f(x)dx = \int_E f_1(x)dx - \int_E f_2(x)dx, \tag{3}$$

where $\quad f_1(x) = f(x), \text{ if } f(x) \geqq 0, \quad f_2(x) = f(x), \text{ if } f(x) < 0,$

$\qquad\qquad = 0, \text{ if } f(x) < 0, \qquad\qquad = 0, \text{ if } f(x) \geqq 0.$

Of course, it is assumed that the functions $f_1(x)$, $f_2(x)$ thus defined are measurable and their integrals exist in accordance with the foregoing definition.

If the function $f(x)$ satisfies the above definition, it is said to be **summable** on the set of points $E$. This term is sometimes applied also to bounded functions satisfying the conditions of a Lebesgue integral. Summability thus plays the same rôle in the Lebesgue theory of integration as integrability does in the Riemann theory. The notion of summability can be extended to the case where $f(x)$ is defined for non-bounded sets $E$.

As a consequence of the foregoing definition of summability, it follows that if $f(x)$ is summable, then $| f(x) |$ is also, and conversely if $| f(x) |$ is summable it follows that $f(x)$ is summable. Moreover, any summable function can be written as the difference of two non-negative functions $f_1(x)$, $f_2(x)$ which are summable in the Lebesgue sense. Hence, in the discussions to follow we may, without loss of generality, assume $f_k(x) \geqq 0$, since all other cases may be made to depend upon this one.

Some of the special properties of summable functions are given in the following theorems.

THEOREM I. *If $f(x)$ is summable on the set $E$, then the subset of $E$ where $f(x)$ becomes infinite has the measure zero.*

From the foregoing discussion, $f_k(x)$ is bounded and measurable. Moreover, we have from the definition of a summable function

$$\int_E f(x)dx = \underset{k\to\infty}{L} \int_E f_k(x)dx = A, \tag{4}$$

where $A$ is a finite number. Denote by $E_1$ the subset of $E$ where $f(x) \leqq k$ and by $E_2$ the subset where $f(x) > k$. We have then

$$E = E_1 + E_2.$$

From (4), together with the definition of a Lebesgue integral, we have

$$\begin{aligned} A &= \underset{k\to\infty}{L} \left\{ \int_{E_1} f_k(x)dx + \int_{E_2} f_k(x)dx \right\} \\ &= \underset{k\to\infty}{L} \left\{ \int_{E_1} f_k(x)dx + k\,m(E_2) \right\}. \end{aligned}$$

Since $k$ and $f_k(x)$ are both positive, it follows that

$$A \geqq k\,m(E_2).$$

The value of $k$ is not bounded, and if it is allowed to increase indefinitely the value of $m(E_2)$ must become arbitrarily small. We may then write with the proper selection of $k$

$$m(E_2) < \epsilon,$$

and consequently in the limit the measure of $E_2$ is zero as $k$ becomes infinite.

THEOREM II. *If $f(x)$ is summable on the point set $E$, then it is summable on any measurable subset of $E$.*

Let $E_1$ be any measurable subset of $E$. We have by hypothesis the condition that

$$\int_E f(x)dx = \underset{k\to\infty}{L} \int_E f_k(x)dx,$$

to prove that the limit

$$\underset{k\to\infty}{L} \int_{E_1} f_k(x)dx \tag{5}$$

exists, where $f_k(x)$ is defined as in the definition of a summable function and is assumed to be non-negative.

Since $E_1$ is a subset of $E$, it follows that $f_k(x)$ defined for the set $E_1$ can not exceed, for any value of $x$, the value of $f_k(x)$ defined for the point set $E$. As $E_1$ is by hypothesis measurable, it follows from the definition of $f_k(x)$ that it is a bounded, measurable function on the set $E_1$. The integral $\int_{E_1} f_k(x)dx$ then exists and

$$\int_{E_1} f_k(x)dx \leqq \int_E f_k(x)dx.$$

Both integrals represent monotone functions of $k$ and the second converges by hypothesis to a finite limit as $k$ increases indefinitely. Consequently, the first integral is bounded and monotone and must therefore converge to a finite limit as $k$ increases without limit; that is, the limit (5) must exist as the theorem requires.

THEOREM III. *Suppose $f(x)$ to be summable on the set of points $E$ and let $E_1$ be any measurable subset of $E$. The integral $\int_{E_1} f(x)dx$ tends to zero with the measure of $E_1$.*

It is sufficient to consider the case where $f(x)$ is non-bounded and non-negative.

From the definition of summability, we may then take $k$ sufficiently large so that

$$\int_E f(x)dx < \int_E f_k(x)dx + \epsilon,$$

where $\epsilon$ is an arbitrarily small positive number. Since $E_1$ is a measurable subset of $E$, it follows from this relation that

$$\int_{E_1} f(x)dx < \int_{E_1} f_k(x)dx + \epsilon. \tag{6}$$

The value of $f_k(x)$ never exceeds $k$; therefore, if we have

$$m(E_1) < \frac{\epsilon}{k},$$

it follows that the integral $\int_{E_1} f_k(x)dx$ is less than $\epsilon$, and hence from (6) we have

$$\int_{E_1} f(x)dx < 2\epsilon,$$

where $2\epsilon$ is arbitrarily small. In other words, the integral $\int_{E_1} f(x)dx$ tends toward zero as $\epsilon$ and therefore $m(E_1)$ approaches zero.

The foregoing theorem is equivalent to saying that the integral of $f(x)$ is an absolutely continuous function of $E_1$.

It is of interest to note how many of the properties of Lebesgue integrals discussed in Art. 51 can be extended so as to apply to any summable function. It becomes at once evident from the definition of summable functions that Theorems I–VI can be so extended by merely substituting in those theorems the term "summable" for "bounded and measurable." The proofs depend upon the use of the laws of limits and may be supplied by the reader.

Theorem VII of Art. 51 does not permit of a corresponding extension without some modification. We shall introduce the notion of **equi-absolute continuity**[1] of the integrals $\int f_n(x)dx$, where $\{f_n(x)\}$ is an infinite sequence of summable functions. These integrals are said to be equi-absolutely continuous on the set $E$, if to every positive $\epsilon$ there corresponds a positive number $\delta$ such that for any $n$ we have

$$\left|\int_e f_n(x)dx\right| < \epsilon,$$

where $e$ is a measurable subset of $E$ having a measure less than $\delta$.

It is important to notice that in case the integrals $\int f_n(x)dx$ are equi-absolutely continuous on the bounded set $E$, then the integrals $\int_E f_n(x)dx$ are bounded for all values of $n$. In fact, $E$ may be divided into a finite number of sets each having a measure less than $\delta$ and the integral on $E$ may be divided into the same finite number of integrals each in absolute value less than $\epsilon$.

With the aid of this property, we may state the following theorem.

THEOREM IV. *Let* $\{f_n(x)\}$ *be a sequence of summable functions, bounded or not, having* $f(x)$ *as its limiting function. If the integrals* $\int f_n(x)dx$ *are equi-absolutely continuous on* $E$, *then* $f(x)$ *is summable on* $E$.

It is sufficient to demonstrate the theorem for the sequence $\{|f_n(x)|\}$. For, to say that the integrals of the given sequence are equi-absolutely continuous is equivalent to saying that those of the sequence $\{|f_n(x)|\}$ are also, since we can always consider only those subsets $E_1$ for which $f(x)$ does not change sign. Moreover, if the sequence $\{f_n(x)\}$ approaches $f(x)$, then the sequence $\{|f_n(x)|\}$ at the same time has the limiting function $|f(x)|$, and consequently $f(x)$ is summable whenever $|f(x)|$ is summable.

---

[1] See Vitali, *Rendiconti del Circolo matematico di Palermo*, Vol. 23 (1907), pp. 137–55. Also de la Vallée Poussin, *Trans. Amer. Math. Soc.*, Vol. 16 (1915), pp. 435–501.

In accordance with the foregoing statements, let us assume that both $f_n(x)$ and $f(x)$ are either positive or zero. If $k$ is any finite number, then we define $[f_n(x)]_k$ to be equal to $f_n(x)$ whenever $f_n(x)$ is less than $k$ and equal to $k$ for those values of $x$ for which $f_n(x) \geqq k$. Let $f_k(x)$ be defined in the same manner. Then both $[f_n(x)]_k$ and $f_k(x)$ are bounded and $[f_n(x)]_k$ approaches $f_k(x)$ as a limiting function. From Theorem VII of Art. 51, we have

$$\int_E f_k(x)dx = \underset{n\to\infty}{L} \int_E [f_n(x)]_k dx.$$

Since $[f_n(x)]_k \leqq f_n(x)$, we have

$$\int_E [f_n(x)]_k \leqq \int_E f_n(x)dx.$$

However, the integrals $\int_E f_n(x)dx$ are bounded as a consequence of the hypothesis that the integrals $\int f_n(x)dx$ are equi-absolutely continuous on $E$. There must then exist a finite number $M$ such that for all values of $k$ we have

$$\int_E f_k(x)dx \leqq M.$$

The integrals $\int_E f_k(x)dx$ are monotone increasing functions of $k$. Since these integrals are bounded, it follows that the limit

$$\underset{k\to\infty}{L} \int_E f_k(x)dx$$

must exist, and hence the function $f(x)$ is summable.

THEOREM V. *Let $\{f_n(x)\}$ be a sequence of summable functions, bounded or not, having $f(x)$ as its limiting function. If the integrals $\int f_n(x)dx$ are equi-absolutely continuous on the set $E$, then*

$$\underset{n\to\infty}{L} \int_E f_n(x)dx = \int_E f(x)dx.$$

From Theorem IV, it follows that $f(x)$ is summable, and hence it must be finite except for a set of points of measure zero. Moreover, the integral of $f(x)$ is equi-absolutely continuous, by virtue of Theorem III.

The value of the integral of any of the functions

$$f_1(x),\ f_2(x),\ \cdots,\ f_n(x),\ \cdots,\ f(x) \tag{7}$$

is not affected by omitting from $E$ those points where the function becomes infinite, because such points form at most a set of measure zero. For the purposes of our demonstration, we may therefore assume that these functions are finite.

Set up the function $\phi_n(x) \equiv |f_n(x) - f(x)|$. As $n$ becomes infinite, $\phi_n(x)$ has the limiting value zero. Moreover, the integrals $\int \phi_n(x)dx$ are equi-absolutely continuous on the set $E$, since that is true of the integral $\int f(x)dx$ and by hypothesis of the integrals $\int f_n(x)dx$. The theorem will be established if we can show that

$$\underset{n\to\infty}{L} \int_E \phi_n(x)dx = 0.$$

Let $M_1, M_2, M_3, \cdots, M_k, \cdots$ be an increasing sequence of positive numbers having no finite upper bound. Denote by $E_k$ the set of those points where at least one of the functions $\phi_n(x)$ exceeds in absolute value the number $M_k$. It follows that $m(E_k)$ approaches zero as $k$ becomes infinite, since this measure is that of the set of points common to the sequence $\{E_k\}$ for all values of $k$, and there are no such common points.

We may write for all values of $n$

$$\int_E \phi_n(x)dx = \int_{E-E_k} \phi_n(x)dx + \int_{E_k} \phi_n(x)dx. \tag{8}$$

We have

$$\underset{n\to\infty}{L} \int_{E-E_k} \phi_n(x)dx = 0;$$

because, for every arbitrarily chosen $\epsilon$ there must exist a $k$ such that for the corresponding point set $(E - E_k)$ we have for all values of $n$ sufficiently large

$$\left| \int_{E-E_k} \phi_n(x)dx \right| < \epsilon.$$

We may also write

$$\underset{n\to\infty}{L} \int_{E_k} \phi_n(x)dx = 0;$$

for, these integrals are equi-absolutely continuous and hence they must approach zero with $m(E_k)$, which, as we have seen, has the limit zero. Hence, from (8) we have the desired result upon passing to the limit, namely,

$$\underset{n\to\infty}{L} \int_E \phi_n(x)dx = 0.$$

THEOREM VI. *Let* $\{f_n(x)\}$ *be a sequence of summable functions having* $f(x)$ *as its limiting function. If we have on the point set* $E$

$$|f_n(x)| < \phi(x),\ n = 1, 2, 3, \cdot \cdot \cdot,$$

*where* $\phi(x)$ *is a positive, summable function, then* $f(x)$ *is summable and*

$$\underset{n\to\infty}{L} \int_E f_n(x)dx = \int_E f(x)dx.$$

As before, we shall prove the theorem for the case where $f_n(x)$ is non-negative. The result can then be extended to the general case since all such cases can be expressed as the difference of two non-negative functions.

By hypothesis we have

$$|f_n(x)| < \phi(x),\ n = 1, 2, 3, \cdot \cdot \cdot, \tag{9}$$

where $\phi(x)$ is summable on the given set $E$. From the summability of $\phi(x)$ it follows that

$$\int_e \phi(x)dx < \epsilon, \qquad \text{(By Theorem III)}$$

where $e$ is a suitably chosen, measurable subset of $E$. From (9) it follows that

$$\left| \int_e f_n(x)dx \right| < \epsilon;$$

that is, the integrals $\int f_n(x)dx$ are equi-absolutely continuous. It follows then from Theorem V that

$$\underset{n\to\infty}{L} \int_E f_n(x)dx = \int_E f(x)dx.$$

THEOREM VII. *Let* $\{f_n(x)\}$ *be a non-decreasing sequence of positive functions having the limiting function* $f(x)$, *which may be finite or infinite. Then we have*

$$\underset{n\to\infty}{L} \int_E f_n(x)dx = \int_E f(x)dx. \tag{a}$$

*However, if* $f(x)$ *is not summable, then both members of this equation become infinite.*

If $f(x)$ is summable, this theorem reduces to the preceding theorem; for, all we need to do is to replace $f(x)$ by $\phi(x)$ as used in that theorem.

If $f(x)$ is not summable, define $f_k(x)$ to be equal to $f(x)$ wherever $f(x)$ is less than $k$, $k$ being any finite number, and equal to $k$ for those points of $E$ where $f(x)$ is equal to or exceeds $k$. Let $[f_n(x)]_k$ be defined in a similar manner. Then just as in the demonstration of Theorem IV, we have

$$\int_E f_k(x)dx = \underset{n\to\infty}{L} \int_E [f_n(x)]_k dx \leq \underset{n\to\infty}{L} \int_E f_n(x)dx.$$

The first integral in this inequality becomes infinite with $k$. As $\{f_n(x)\}$ is non-decreasing and independent of $k$, it follows that $\underset{n\to\infty}{L} \int_E f_n(x)dx$ becomes infinite as the theorem states.

THEOREM VIII. *Let* $\{f_n(x)\}$ *be a sequence of non-negative, summable functions having the limiting function* $f(x)$. *A necessary and sufficient condition that* $f(x)$ *is summable and also that*

$$\int_E f(x)dx = \underset{n\to\infty}{L} \int_E f_n(x)dx \tag{10}$$

*is that the integrals* $\int f_n(x)dx$ *are equi-absolutely continuous on the point set* $E$.[1]

We have already seen that the given condition is sufficient; that is, if the integrals $\int f_n(x)dx$ are equi-absolutely continuous, then $f(x)$ is summable and equation (10) follows. It remains to show that the condition is necessary.

Assuming that $f(x)$ is summable and that equation (10) is valid, we must show that the integrals $\int f_n(x)dx$ are equi-absolutely continuous on $E$. We shall do this by showing that if the integrals $\int f_n(x)dx$ are not equi-absolutely continuous, then the relation expressed by (10) is not valid.

Let $M_1, M_2, \cdot\cdot\cdot, M_k, \cdot\cdot\cdot$ be a sequence of positive numbers which increase without limit and let $E_k$ be that subset of $E$ where at least one of the functions $f_n(x)$ exceeds $M_k$. It follows that $m(E_k)$ approaches zero as $k$ becomes infinite since it is the measure of a set of points common to the sets of the sequence $\{E_k\}$; that is, of a set of points where $f(x)$ becomes infinite.

As it is assumed that the integrals $\int f_n(x)dx$ are not equi-absolutely continuous on $E$, it follows that we may select an $\epsilon$ such that however small the positive number $\delta$ may be chosen we have

$$\int_e f_n(x)dx > \epsilon, \tag{11}$$

---

[1] See de la Vallée Poussin, *Trans. Amer. Math. Soc.*, Vol. 16 (1915), p. 447.

where $e$ is a subset of $E$ and of measure less than $\delta$. It is to be observed that the integrals $\int f_n(x)dx$ are, however, equi-absolutely continuous on the complementary set $C(E_k)$ where the functions $f_n(x)$ are bounded. We may then conclude that for all values of $k$ and for a sufficiently large value $n$ we have

$$\int_{E_k} f_n(x)dx > \epsilon.$$

Because the integrals $\int f_n(x)dx$ are equi-absolutely continuous on the point set $C(E_k)$, it follows from Theorem V, that

$$\underset{n\to\infty}{L} \int_{C(E_k)} f_n(x)dx = \int_{C(E_k)} f(x)dx. \tag{12}$$

Consequently, we have

$$\underset{n\to\infty}{\overline{L}} \int_E f_n(x)dx = \underset{n\to\infty}{L} \int_{C(E_k)} f(x)dx + \underset{n\to\infty}{\overline{L}} \int_{E_k} f_n(x)dx$$
$$> \int_{C(E_k)} f(x)dx + \epsilon.$$

Upon passing to the limit as $k$ becomes infinite, we have

$$\underset{n\to\infty}{L} \int_E f_n(x)dx > \int_E f(x)dx + \epsilon.$$

This conclusion contradicts the relation expressed in (10). Hence the assumption that the integrals $\int f_n(x)dx$ are not equi-absolutely continuous on $E$ is not valid, and from this contradiction the given theorem follows.

**54. Indefinite Lebesgue integrals.** As with Riemann integrals, we may define the indefinite integral $\phi(x)$ by the relation

$$\phi(x) \equiv \int_a^x f(x)dx, \quad a \leqq x \leqq b, \tag{1}$$

where we assume $f(x)$ to be a single-valued, summable function, but not necessarily bounded in the interval $(a, b)$. As we shall see, the value of $\phi(x)$ is then determined except as to an additive constant.

Lebesgue[1] has generalized the notion of the indefinite integral by replacing the subinterval $(a, x)$ by any measurable subset of $E_1$

---

[1] See *Encyclop. der Math. Wiss.*, Bd. II$_3$, p. 113; also Lebesgue, *Ann. Ecol. Norm.* (3), Vol. 27 (1910), pp. 361–450.

of the given point set $E$ over which the definite integral is taken. Thus, we may put

$$\phi(E_1) = \int_{E_1} f(x)dx,$$

where $\phi(E_1)$ is defined as the generalized indefinite integral.

Some of the more important properties of indefinite integrals are stated in the following theorems.

THEOREM I. *Let $f(x)$ be a summable function in the interval $(a, b)$. Then the indefinite integral*

$$\phi(x) \equiv \int^x f(x)dx, \quad a \leqq x \leqq b,$$

*is an absolutely continuous function.*

Let $\{I_k\}$ be any finite or enumerably infinite set of subintervals lying in the interval $(a, b)$, made by the insertion of a set of intermediate points $\{x_k\}$ between $a$ and $b$. We may then write

$$I_k \equiv x_k - x_{k-1},\ \Sigma I_k \equiv I,\ \Delta_k\phi(x) \equiv \int_{I_k} f(x)dx.$$

Making use of this notation, we have

$$\Sigma\,|\,\Delta_k\phi(x)\,| = \Sigma\,\left|\int_{I_k} f(x)dx\right| \leqq \int_I |\,f(x)\,|\,dx. \tag{2}$$

By Theorem III, Art. 53, this last integral, however, approaches zero with the measure of the set $\{I_k\}$, that is with the measure of $I$. The first term is the sum of the absolute values of the differences of $\phi(x)$ in the set of intervals $\{I_k\}$; that is,

$$\Sigma\,|\,\Delta_k\phi(x)\,| = |\,\phi(x_1) - \phi(x)\,| + |\,\phi(x_2) - \phi(x_1)\,| + \cdot\ \cdot\ \cdot,$$

where the points of the set $\{x_k\}$ are arbitrarily chosen in the interval $(a, b)$. It follows from (2) that the sum of the oscillations of $\phi(x)$ in any set of intervals $\{I_k\}$ in $(a, b)$ approaches zero with the sum of the lengths of these intervals. This is the condition that the indefinite integral $\phi(x)$ is absolutely continuous in $(a, b)$.

COR. *If $f(x)$ is summable in the interval $(a, b)$, the indefinite integral*

$$\phi(x) \equiv \int_a^x f(x)dx$$

*is a continuous function.*

This proposition follows as a special case of the foregoing theorem. For, if $\phi(x)$ is absolutely continuous, it is necessarily continuous.

THEOREM II. *Let $f(x)$ be a summable function in the interval $(a, b)$. Then the indefinite integral*

$$\phi(x) \equiv \int_a^x f(x)dx, \quad a \leqq x \leqq b, \tag{3}$$

*is of limitied variation and its total limited variation in $(a, b)$ is*

$$\int_a^b | f(x) | dx.$$

That $\phi(x)$ is of limited variation follows from Theorem I; for, being absolutely continuous, it follows from Theorem VII of Art. 38 that it must also be of limited variation.

We shall now find an expression for the total variation of $\phi(x)$ in the interval $(a, b)$. Divide the interval $(a, b)$ in any manner into consecutive intervals. Denote by $(\alpha)$ those subintervals where the integral $\int_\alpha f(x)dx$ is positive and by $\beta$ those where the integral $\int_\beta f(x)dx$ is negative. Using the notation of Art. 38, we may write

$$p = \Sigma \int_\alpha f(x)dx, \quad n = -\Sigma \int_\beta f(x)dx,$$

and the variation $V$ of $\phi(x)$ is then given by

$$V = p + n.$$

If we denote by $E_p$ the set of points where $f(x) \geqq 0$ and by $E_n$ the set where $f(x) < 0$, then if, as in Art. 38, $P$ and $N$ denote the least upper bounds of $p$ and $n$, respectively, for all possible methods of subdivision of $(a, b)$, we have

$$P \leqq \int_{E_p} f(x)dx, \quad N \leqq -\int_{E_n} f(x)dx. \tag{14}$$

We shall now show that $\int_{E_p} f(x)dx$ can not exceed $P$. A similar argument can be readily formulated by the reader to show that $-\int_{E_n} f(x)dx$ can not exceed $N$.

As $E_p$ is measurable, it can be expressed as the sum of a closed set and one whose exterior measure is less than an arbitrarily small number $\epsilon$. Hence, by the aid of the generalized form of the Heine-Borel Theorem applied to the closed subset of $E_p$, we can enclose the points of $E_p$ in a set of intervals $\delta_1$ whose measure can be made to differ from that of $E_p$ by as little as we please. We may therefore write

$$\int_{\delta_1} f(x)dx = \int_{E_p} f(x)dx + \int_{\delta_2} f(x)dx,$$

where $m(\delta_2)$ and hence the integral $\int_{\delta_2} f(x)dx$ is arbitrarily small. As the integral in the first member of this equation expresses a value of $p$ for the intervals $\delta_1$, it follows that $\int_{E_p} f(x)dx$ can not exceed the least upper bound of all values of $p$; namely, $P$.
Hence, we obtain by aid of (4)

$$V_{ab} = P_{ab} + N_{ab} = \int_{E_p} f(x)dx - \int_{E_n} f(x)dx$$

$$= \int_a^b | f(x) | dx.$$

THEOREM[1] III. *If one of the derivative numbers of a continuous function* $f(x)$, *say* $D^+f(x)$, *is bounded in the interval* $(a, b)$, *we have*

$$f(x) - f(a) = \int_a^x D^+f(x)dx, \qquad a \leqq x \leqq b,$$

*and* $f(x)$ *has a finite derivative at all points of* $(a, b)$ *with the exception of at most a set of measure zero.*

The derivative number $D^+f(x)$ is bounded and by Cor. 1 to Theorem XII, Art. 49, it is also measurable, and hence it is summable.

Let the indefinite interval $(-\infty, \infty)$ on the axis of ordinates be divided into subintervals $(l_n, l_{n+1})$ such that we have

$$l_{n+1} - l_n < \epsilon$$

as $n$ takes all positive and negative integral values and zero, and let $l_0 = 0$. Denote by $e_n$ the points of $(a, b)$ where $l_n < D^+f(x) \leqq l_{n+1}$ and arrange the sets $e_n$ in the order

$$e_0, e_1, e_{-1}, e_2, e_{-2}, \cdot \cdot \cdot , e_n, e_{-n}, \cdot \cdot \cdot .$$

Select a set of positive numbers $a_n$ less than 1, such that

$$a_0 + a_1 | l_1 | + a_{-1} | l_{-1} | + \cdot \cdot \cdot + a_n | l_n | + a_{-n} | l_{-n} | + \cdot \cdot \cdot$$

is less than $\epsilon$.

Inclose $e_0$ in a set of non-overlapping intervals $A_0$, and inclose the complementary set $C(e_0)$ in a set of intervals $I_0$ chosen in such a manner that the measure of the parts common to $A_0$ and $I_0$ shall not exceed $a_0$. Likewise inclose $e_1$ in a set of intervals $A_1$ and

---

[1] See Lebesgue's *Leçons sur l'intégration*, pp. 120–3; also *Atti dei Lincei Rendiconti*, (5) Vol. 15 I, pp. 433–8, 551–8, 674–84; Vol. 15 II pp. 3–8, 358–68; Vol. 16 I, pp. 92–100, 283–90.

$C(e_0+e_1)$ in a set of intervals $I_1$, such that the common parts of these intervals shall have a measure not greater than $a_1$ and both $A_1$ and $I_1$ shall lie within $I_0$. Proceeding in this manner, we inclose $e_n$, where $n$ may be either positive or negative, in a set of intervals $A_n$ and $C(e_0+e_1+\cdots+e_n)$ in a set of intervals $I_n$ such that the measure of the parts common to the two sets shall not exceed $a_n$ and both $A_n$ and $I_n$ shall lie within $I_{n-1}$.

It follows that $\quad m(A_n) - m(e_n) \leqq a_n,$

and $A_n$ has in common with all other $A$'s a set of intervals whose measure does not exceed

$$a_0 + a_1 + \cdots + a_n \leqq N_n. \tag{5}$$

Since $\quad \Sigma \mid l_n \mid m(A_n) - \Sigma \mid l_n \mid m(e_n) \leqq \Sigma \mid l_n \mid a_n < \epsilon, \qquad (6)$

we see that $\Sigma \mid l_n \mid m(A_n)$ and $\Sigma \mid l_n \mid m(e_n)$ are both convergent or both divergent, and in the former case the difference is less than $\epsilon$. For each $e_n$, we have

$$\mid l_n \mid - \epsilon < D^+ f(x) < \mid l_n \mid + \epsilon,$$

and hence

$$\Sigma \mid l_n \mid m(e_n) - \epsilon(b-a) < \int_a^b D^+ f(x)dx < \Sigma \mid l_n \mid m(e_n) + \epsilon(b-a). \tag{7}$$

Each $x$ in $(a, b)$ is a point in some $e_n$. Let $A_n'$ be that interval of the set $A_n$ which contains $x$. Let $(x, x + \Delta x)$ be the longest interval, of length not greater than $\epsilon$, to the right of $x$ contained in $A_n'$ and satisfying the condition

$$\mid l_n \mid - \epsilon \leqq \frac{f(x+\Delta x) - f(x)}{\Delta x} \leqq \mid l_{n+1} \mid + \epsilon,$$

whence

$$\Delta x\{\mid l_n \mid - \epsilon\} \leqq f(x+\Delta x) - f(x) \leqq \Delta x\{\mid l_n \mid + \epsilon\}.$$

We may cover the interval $(a, b)$ by a chain of intervals $(x, x+\Delta x)$ beginning with $a$. For the left-hand end-points $x$ which belong to the same $e_n$, let $B_n$ be the intervals of the corresponding part of the chain. For that part of the chain for which $n \leqq N$, where $N$ is finite, we have

$$\sum_{n=1}^{N} \mid l_n \mid m(B_n) - \epsilon(b-a) \leqq \sum_{n=1}^{N} \{f(x+\Delta x) - f(x)\}$$
$$\leqq \sum_{n=1}^{N} \mid l_n \mid m(B_n) + \epsilon(b-a). \tag{8}$$

The points of $A_n$ which are not a part of $B_n$ form a part of a set $A_{n+p}$ $(p \neq 0)$, and their measure can not exceed $N_n$. Hence, we have

$$\sum_{n=1}^{N} | l_n | m(A_n) - \sum_{n=1}^{N} | l_n | m(B_n) \leqq \sum_{n=1}^{N} | l_n | N_n.$$

The values of $a_n$ can be so chosen that

$$\sum_{n=1}^{N} | l_n | N_n < \epsilon,$$

whatever value is given to $N$. We then have

$$\sum_{n=1}^{N} | l_n | m(A_n) - \epsilon < \sum_{n=1}^{N} | l_n | m(B_n) < \sum_{n=1}^{N} | l_n | m(B_n) + \epsilon. \quad (9)$$

For all values of $n$, we have

$$m(A_n) \geqq m(e_n) \geqq m(B_n),$$

and hence obtain from (8)

$$\sum_{n=1}^{N} | l_n | m(e_n) - \epsilon(b - a + 1) < \sum_{n=1}^{N} \{f(x + \Delta x) - f(x)\}$$
$$< \sum_{n=1}^{N} | l_n | m(e_n) + \epsilon(b - a + 1).$$

For values of $N$ sufficiently large, the difference between $\sum_{n=1}^{N} | l_n | m(e_n)$ and $\sum_{n=1}^{\infty} | l_n | m(e_n)$ is less than $\epsilon$. Hence, we have

$$\Sigma | l_n | m(e_n) - \epsilon(b - a + 2) < \Sigma \{f(x + \Delta x) - f(x)\}$$
$$< \Sigma | l_n | m(e_n) + \epsilon(b - a + 2). \quad (10)$$

By comparing (7) and (10), it follows that $f(b) - f(a)$ lies between the values

$$\int_a^b D^+f(x)dx - \epsilon(b - a + 3), \int_a^b D^+f(x)dx + \epsilon(b - a + 3).$$

Since this relation holds for every choice of $\epsilon$ however small, it follows that for $a \leqq x \leqq b$, we have

$$f(x) - f(a) = \int_a^x D^+f(x)dx.$$

A similar proof shows this relation to hold for each of the other derivative numbers. We may then write

$$f(x) - f(a) = \int_a^x D^+f(x)dx = \int_a^x D_+f(x)dx = \int_a^x D^-f(x)dx$$
$$= \int_a^x D_-f(x)dx.$$

Because these integrals all exist and are equal, the integrands must be equal at all points of $(a, b)$ with the exception of at most a set of measure zero. Hence, $f(x)$ must have a finite derivative at all points of $(a, b)$, except for a set of at most measure zero, as the theorem requires.

COR. 1. *Every bounded derivative is summable.*

This corollary follows as an immediate consequence of the foregoing theorem; for, if $f(x)$ has a bounded derivative at every point of $(a, b)$, then it is necessarily continuous and the conditions of the theorem are satisfied.

COR. 2. *Every bounded derivative number of a continuous function $f(x)$ is summable.*

As we have seen, if $f(x)$ has a bounded derivative number, then all the derivative numbers of $f(x)$ are bounded and the corollary follows from the foregoing theorem.

THEOREM IV. *Let $f(x)$ be a summable function in the interval $(a, b)$. The indefinite Lebesgue integral*

$$\phi(x) \equiv \int_a^x f(x)dx$$

*has $f(x)$ as its derivative at all points of $(a, b)$, except at most at a set of points of measure zero.*[1]

Let us first consider the case of a measurable function $f(x)$ which takes only the values zero and one. Let $\phi(x)$ be the corresponding indefinite integral. Let $E$ be the set of points for which $f(x)$ takes the value one. Inclose the points of $E$ in a set of intervals $A_k$, the sum of whose lengths is $m(E) + \epsilon_k$, where $\epsilon_k$ approaches zero as $k$ increases indefinitely. Denote by $E_1$ the set of points common to the sets of intervals

$$A_1, A_2, \cdots, A_k, \cdots.$$

Then $E$ is contained in $E_1$ and differs from it by a set of points of measure zero. In the calculation of $\phi(x)$, we can replace $f(x)$ by another function $f_0(x)$ of such a character that it takes the value one at the set of points $E_1$. Let $f_k(x)$ be a function which takes the value one at all points of $A_k$, being zero at all other points. Then $f_0(x)$ is the limit of the decreasing functions $f_k(x)$ as $k$ becomes infinite. Denote the indefinite integral of $f_k(x)$ by $\phi_k(x)$. For a positive increment of $x$, the increment of $\phi_k(x)$ is at least equal to that of $\phi(x)$. It then follows from the definition of a derivative number that

---

[1] See Lebesgue, *Leçons sur l'intégration*, p. 124.

$$D\phi_k(x) \geqq D\phi(x) \geqq 0, \tag{11}$$

where $D$ denotes any one of the four derivative numbers.

To compute the value of $D\phi_k(x)$ within $A_k$, we have from the definition of $\phi_k(x)$

$$\begin{aligned}\frac{\phi_k(x + \Delta x) - \phi_k(x)}{\Delta x} &= \frac{1}{\Delta x}\left[\int_a^{x+\Delta x} f_k(x)dx - \int_a^x f_k(x)dx\right]\\ &= \frac{1}{\Delta x}\left[\int_x^{x+\Delta x} f_k(x)dx\right].\end{aligned}$$

Since $f_k(x)$ is one for all values of $x$ within $A_k$, this integral reduces to $\Delta x$, and we have for all such values of $x$, irrespective of the values of $\Delta x$,

$$\frac{\phi_k(x + \Delta x) - \phi_k(x)}{\Delta x} = 1.$$

Hence, within this interval $D\phi_k(x)$ is equal to one for each value of $x$; that is, at all inner points of $A_k$. Consequently, we may conclude that $D\phi_k(x)$ is different from zero at the inner points of $A_k$ and at other points which, as $k$ becomes infinite, form a set of measure zero. From (11) it follows that $D\phi(x)$ can be different from zero only at the points of $E_1$ (or of $E$) and otherwise at a set of points having the measure zero.

The values of $D\phi(x)$ are bounded since by (11) they can never exceed the values of $D\phi_k(x)$ which are at most equal to one. Moreover, $D\phi(x)$ is measurable since $\phi(x)$ is continuous by the corollary to Theorem I. It is therefore summable. The conditions of Theorem III are satisfied, and we have

$$\int_a^b D\phi(x)\,dx = \phi(b) - \phi(a), \tag{12}$$

Since $f(x)$ is zero or one we have

$$\phi(b) - \phi(a) = m(E). \tag{13}$$

From (12) and (13) we may conclude that $D\phi(x)$ is equal to one for all points of $E$ with the exception of at most a set of points of measure zero. These relations hold for all the derivative numbers, and hence we can say that $f(x)$ is the derivative of $\phi(x)$ with the exception of at most a set of points of measure zero.

We shall now consider the case in which $f(x)$ is a bounded, measurable function. Denote by $L$ and $l$ the least upper and the greatest lower bounds respectively of $f(x)$ in the interval $(a, b)$. These two values, if laid off on the $Y$-axis, give the limits of variation of $f(x)$ in $(a, b)$. Divide this interval into partial intervals by inserting the following points:

$$y_0(= l),\ y_1,\ y_2,\ \cdots,\ y_{k-1},\ y_k,\ \cdots,\ y_n(= L).$$

We shall so select the intermediate points that for all values of $k$ we have

$$y_k - y_{k-1} < \epsilon,$$

where $\epsilon$ is an arbitrarily chosen positive number. Let $\psi_k(x)$ $(k = 1, 2, 3, \cdots n)$ take the value one if $x$ is such a value that $y_k \leqq f(x) < y_{k+1}$ and suppose $\psi_k(x)$ to be zero for all other values of $x$ in $(a, b)$. Denote by $\Psi_k(x)$ $(k = 1, 2, 3, \cdots n)$ a function which takes the value one if $x$ is contained in a set of points for which $y_{k-1} < f(x) \leqq y_k$. For all other values, suppose $\Psi_k(x)$ to be zero. Put

$$\psi(x) \equiv \sum_{k=0}^{n} y_k\psi_k(x),\quad \Psi(x) \equiv \sum_{k=0}^{n} y_k\Psi_k(x).$$

As $\psi_k(x)$ and $\Psi_k(x)$ take only the values one and zero, then from their definition it follows that, as $n$ increases indefinitely in such a manner that $\epsilon$ approaches zero, the functions $\psi(x)$, $\Psi(x)$ approach uniformly the function $f(x)$ as a limiting function. Upon integrating these functions, we have

$$\int_a^x \psi(x)dx \leqq \int_a^x f(x)dx \leqq \int_a^x \Psi(x)dx,\ a \leqq x \leqq b,$$

where the first and last integrals approach a common value as $\epsilon$ approaches zero. This common value is equal to the $\int_a^x f(x)dx$. However, from the discussion in the first part of this proof we know that the functions $\psi(x)$, $\Psi(x)$ are the derivatives of their indefinite integrals, except for a set of points at most of measure zero. Consequently, as $\epsilon\,(> y_k - y_{k-1})$ approaches zero, the derivative numbers of $\phi(x)$ are at least equal to $f(x)$, except at a set of points of measure zero; for, in every interval the increment of the integral $\int_a^x f(x)dx$ is at least equal to that of $\int_a^x \psi(x)dx$. Likewise, the derivative numbers of $\phi(x)$ are at most equal to $f(x)$, except for a set of points of measure zero; since the increment of the integral $\int_a^x f(x)dx$ is at most equal to that of $\int_a^x \Psi(x)dx$ in every interval. Hence, as $n$ becomes infinite in such a manner that $\epsilon$ approaches zero, the derivative of the indefinite integral $\phi(x)$ is equal to $f(x)$ for all points of $(a, b)$ with at most the exception of a set of points of measure zero.

Finally, let us consider the case in which $f(x)$ is any summable function. As in the previous discussions, it is sufficient to consider $f(x)$ as positive or zero.

Let
$$f_n(x) = f(x), \text{ if } f(x) \leqq n,$$
$$= n, \text{ if } f(x) > n.$$

Put
$$\phi_n(x) \equiv \int_a^x f_n(x)dx.$$

We have then
$$D^+\phi_n(x) \leqq D^+\phi(x).$$

Since $f_n(x)$ is bounded, it follows from the foregoing discussion that

$$D^+\phi_n(x) = f_n(x)$$

for all values of $n$ and for all values of $x$ in $(a, b)$ with the exception of at most a set of measure zero. Hence, for the same values of $x$, we have

$$D^+\phi(x) \geqq f_n(x).$$

As this relation holds for all values of $n$, we have

$$D^+\phi(x) \geqq \underset{n\to\infty}{L} f_n(x) = f(x).$$

However, as in the foregoing discussion,

$$\int_a^b D^+\phi(x)dx \leqq \int_a^b f(x)dx$$

or
$$\int_a^b [D^+\phi(x) - f(x)]dx \leqq 0.$$

As $D^+\phi(x)$ is never less than $f(x)$, it follows that

$$D^+\phi(x) = f(x)$$

for all values of $x$ with at most a set of measure zero. A similar proof shows this relation to hold for the other derivative numbers, and hence the theorem follows.

COR. *If $f(x)$ is bounded and measurable in the interval $(a, b)$, the indefinite integral*

$$\phi(x) \equiv \int_a^x f(x)dx,$$

*has $f(x)$ as a derivative at all points of $(a, b)$ with the exception of at most a set of measure zero.*

This corollary follows from the foregoing theorem because of the fact that every function $f(x)$ which is bounded and measurable is also summable.

If the given function has a bounded derivative, we have the following theorem.

THEOREM V. *Let $F(x)$ be a function having a bounded derivative $f(x)$, then the Lebesgue integral of $f(x)$ exists and satisfies the relation*

$$\int^b f(x)dx = F(b) - F(a).$$

We have by hypothesis

$$f(x) = \frac{dF}{dx} \equiv \underset{\Delta x \to 0}{L} \frac{F(x + \Delta x) - F(x)}{\Delta x}, \; a \leqq x \leqq b, \; a \leqq x + \Delta x \leqq b. \quad (14)$$

As we know, a derivative function, and hence $f(x)$, is at most pointwise discontinuous and belongs to Baire's class one. It has been shown in Art. 49 that every function of this class is a measurable function. The function $f(x)$ is also bounded by hypothesis. It follows from the theorem of Art. 50 that $f(x)$ is integrable in the Lebesgue sense. As the limit in (14) exists and is equal to $f(x)$, its value could be obtained by taking the limit through any sequence of values of $\Delta x$ dense at $\Delta x = 0$. Consequently, by Theorem VI, Art. 53, since $F(x)$ is continuous and therefore summable, we have

$$\underset{\Delta x \to 0}{L} \int_a^b \frac{F(x + \Delta x) - F(x)}{\Delta x} dx = \int_a^b f(x)dx. \quad (15)$$

But since $\displaystyle\int_a^b F(x + \Delta x)dx = \int_{a+\Delta x}^{b+\Delta x} F(x)dx,$

we may write

$$\int_a^b [F(x + \Delta x) - F(x)]dx = \int_{a+\Delta x}^{b+\Delta x} F(x)dx - \int_a^b F(x)dx$$
$$= \int_b^{b+\Delta x} F(x)dx - \int_a^{a+\Delta x} F(x)dx.$$

Dividing by $\Delta x$ and passing to the limit as $\Delta x$ approaches zero, we have by aid of the Law of the Mean

$$\underset{\Delta x \to 0}{L} \int_a^b \frac{[F(x + \Delta x) - F(x)]dx}{\Delta x} = F(b) - F(a).$$

Therefore, we have the desired result, namely

$$\int_a^b f(x)dx = F(b) - F(a).$$

By comparing the foregoing theorem with the corresponding theorem for Riemann integrals, it will be seen that in the former case we made use of the condition that $f(x)$ was integrable, while in the present case that restriction can be removed, provided $f(x)$ is bounded.

THEOREM VI. *Let* $f(x)$ *be a bounded, monotone function in the interval* $(a, b)$. *Then* $f(x)$ *has a finite derivative at every point of* $(a, b)$ *with the exception of at most a set of points of measure zero.*[1]

We shall consider the theorem for the case in which $f(x)$ is a positive, non-decreasing function. A similar proof holds for other possible cases and is left to the reader to supply.

The properties of the given function are preserved if we add $Mx$ to it, where $M$ is a positive, finite number. The derivative number $D^+F(x)$ of the function

$$y \equiv F(x) = f(x) + Mx$$

is then never less than $M$. Moreover, the function $F(x)$ may be continuous or discontinuous, but it is not a constant throughout any subinterval of $(a, b)$.

The given function $y = f(x)$ can have only ordinary discontinuities. If such a discontinuity occurs at $x_0$, the inverse function

$$x = \psi(y)$$

is defined as the value $x_0$ at all the corresponding values of $y$. Hence, the inverse function is continuous and has a bounded derivative number $D^+\psi(y)$; for, this derivative number does not exceed $\frac{1}{M}\cdot$ The derivative number $D^+\psi(y)$ may be zero for some values of $y$, since $\psi(y)$ may be constant for certain subintervals. It can not be negative, however. It follows from Theorem III that $\psi(y)$ must have a finite derivative at all points except at most a set of measure zero. To this set of $y$-points there corresponds a set of $x$-points also of measure zero. For, since the $y$-points in question form a set of measure zero, they may be inclosed in a set of intervals $\{I_n\}$ the sum of whose lengths is less than an arbitrarily small number $\delta$. Let $\delta_n$ be the length of the interval $I_n$. In each of the intervals $I_n$ the derivative number $D^+\psi(y)$ is bounded, say less than $A$. This value $A$ is also equal to or greater than the least upper bound of the difference ratio. Then the values of $x$ which correspond to the extreme values of $y$ in $I_n$ must differ by less than $\delta_n A$. Consequently the set of $x$-points corresponding to those of $y$ in $\{I_n\}$ can be inclosed in a set of intervals the sum of whose lengths is less than $A \cdot \delta$. But, as $\delta$ is arbitrarily small, the $x$-points in question form a set of measure zero.

---

[1] See W. H. Young, *Quarterly Journal of Pure and Applied Math.* (1910), Vol. 42, p. 79.

We have shown that $\psi(y)$ has a finite derivative at all points with the exception of at most a set of measure zero. It now follows that $F(x)$ has a finite or infinite derivative at all points of $(a, b)$ with at most an exception of a set of points of measure zero; namely the points corresponding to the $y$-points inclosed in the intervals $\{I_n\}$. However, the points where the derivative is infinite form a set of measure zero. For, if the derivative is infinite it is positively infinite. Suppose the measure of the set of points for which this is true is of measure greater than zero, say of measure $N$, where $N > 0$. Then the sum of the lengths of any set of non-overlapping intervals $\{I_n'\}$ inclosing these points must be at least equal to $N$. Because the derivative is infinite, the increment of the function over any of these intervals must exceed $I_n' \cdot B$, where $B$ can be chosen arbitrarily large. Consequently the sum of the increments of the function $F(x)$ in the intervals $\{I_n'\}$ exceeds $N \cdot B$. As $B$ is to be chosen arbitrarily large and $N$ is different from zero, $N \cdot B$ may be made arbitrarily large, which is impossible since $F(x)$ is bounded.

It follows then that $F(x)$, and hence $f(x)$, has a finite derivative at all points of $(a, b)$ with at most the exception of a set of points of measure zero.

THEOREM VII. *Let $f(x)$ be a continuous, monotone, non-decreasing function in $(a, b)$. Then any of the derivative numbers $Df(x)$ is summable in $(a, b)$, and we have*

$$\int_a^b D f(x) dx \leqq f(b) - f(a).$$

Put

$$\begin{aligned} D f_n(x) &= D f(x), \text{ where } D f(x) \leqq n, \\ &= n, \qquad \text{where } D f(x) > n. \end{aligned}$$

The derivative number $D f_n(x)$ is bounded and measurable. Hence, the indefinite integral

$$\phi_n(x) = \int_a^x D f_n(x) dx$$

exists and has a derivative $\phi_n'(x) = D f_n(x)$ at all points of $(a, b)$ with the exception of at most a set $E$ of measure zero. Inclose the points $E$ in a set of non-overlapping intervals $\{I\}$ of measure $\frac{\epsilon}{n}$. Denote by $I_x$ that portion of the intervals $\{I\}$ lying in the interval $(a, x)$.

Consider the function

$$F(x) = f(x) + n \cdot m(I_x).$$

On the set $E$, we have $\quad D\,F(x) \geqq n,$

while on $C(E)$ we have

$$D\,F(x) \geqq D\,f(x) \geqq D\,f_n(x).$$

Consequently, for an increasing $x$ the increment of $F(x)$ is equal to or greater than that of $\phi_n(x)$. For the interval $(a, b)$, we then have

$$f(b) - f(a) + n \cdot m\,\{I\} = f(b) - f(a) + \epsilon \geqq F(b) - F(a) \geqq \int_a^b D\,f_n(x)dx.$$

Therefore, the integral $\int_a^b D\,f_n(x)dx$ is bounded for all values of $n$ and hence must approach a finite limit as $n$ becomes infinite. Denote by $E_n$ the points of $(a, b)$ where $D\,f(x) > n$. We then have

$$\int_a^b D\,f_n(x)dx > n \cdot m(E_n).$$

But since the integral $\int_a^b D\,f_n(x)dx$ approaches a finite limit as $n$ increases indefinitely, it follows that $\underset{n\to\infty}{L}\ m(E_n)$ must be zero. Hence, for all values of $n$, we have

$$\int_{C(E_n)} D\,f(x)dx \leqq \int_a^b D\,f_n(x)dx \leqq f(b) - f(a) + \epsilon.$$

As $f(x)$ is monotone, non-decreasing and $\int_a^b D\,f_n(x)dx$ is bounded, the first integral in the above inequality approaches a finite limit as $n$ increases indefinitely. This limit is the integral $\int_a^b D\,f(x)dx$. We therefore have

$$\int_a^b D\,f(x)dx \leqq f(b) - f(a).$$

THEOREM VIII. *In the interval* $(a, b)$ *let* $f(x)$ *be a function of limited variation. Then* $f(x)$ *has a finite derivative everywhere in* $(a, b)$ *with the exception of at most a set of points of measure zero, and this derivative is summable in* $(a, b)$.[1]

By Theorem II, Art. 38, it follows that a function of limited variation can be expressed as the difference of two bounded, monotone functions both of which are non-decreasing or non-increasing. As we have seen, any such function has a finite derivative at every point of the interval of definition with the possible exception of a set of points of measure zero. Moreover, this derivative is summable. The difference of two such functions would have the same property, and hence the foregoing theorem follows as a consequence.

---

[1] Cf. Lebesgue, *Leçons sur l'intégration*, p. 128. The proof of Lebesgue, however, involves the condition that $f(x)$ is continuous.

COR. *If $f(x)$ is absolutely continuous in the interval $(a, b)$, then it has a finite derivative at every point of $(a, b)$ with at most the exception of a set of points of measure zero.*

This corollary follows from the fact that every absolutely continuous function is also of limited variation and hence the foregoing theorem applies.

THEOREM IX.[1] *A necessary and sufficient condition that $f'(x)$ is summable in the interval $(a, b)$ and that*

$$f(b) - f(a) = \int_a^b f'(x)dx \tag{16}$$

*is that $f(x)$ is absolutely continuous in $(a, b)$.*

The given condition is necessary; for, assuming that $f'(x)$ is summable and that the relation (16) holds, it follows from Theorem I that $\int_a^x f'(x)dx$ defines an absolutely continuous function, and hence $f(x)$ must be absolutely continuous.

The condition of the theorem is also sufficient; that is, if we assume $f(x)$ to be absolutely continuous, we can show that the relation given in (16) follows. From the absolute continuity of $f(x)$ it follows that it is continuous and of limited variation, and hence by Theorem VIII it has a finite derivative $f'(x)$ everywhere in $(a, b)$ with the exception of at most a set $E$ of measure zero. Moreover, this derivative is summable. As $E$ is of measure zero, the points of this set can be regarded as the inner points of an enumerable set of non-overlapping intervals $\{I_n\}$, the sum of whose lengths is arbitrarily small.

---

[1] Lebesgue has established the following theorem, where the given function $f(x)$ is continuous and $f'(x)$ is replaced by a finite derivative number (See *Leçons sur l'intégration*, p. 122).

THEOREM. Let $f(x)$ be a continuous function and let one of its derivative numbers, say $D^{+1}f(x)$, be finite for all values of $x$ in a given interval $(a, b)$. Then a necessary and sufficient condition that $D^{+}f(x)$ is summable in $(a, b)$ is that $f(x)$ is of limited variation.

We shall now define an auxiliary function $f_1(x)$ as follows. In the set $C\{I_n\}$ complementary to the set of intervals $\{I_n\}$, let

$$f_1(x) = f(x),$$

while in the set of intervals $\{I_n\}$ where $f'(x)$ is indeterminate or infinite let $f_1(x)$ take the values along a straight line joining points representing $f(x)$ at the extremities of the interval. Then in the interval $(a, b)$, $f_1'(x)$ is bounded, and hence by Theorem V we have

$$f_1(b) - f_1(a) = \int_a^b f_1'(x)dx.$$

Since $f_1(x)$ and $f(x)$ are identical on the set $C(I_n)$, as well as at the extremities of $(a, b)$ and of any subinterval $I_n$, we have

$$f(b) - f(a) = \int_{C\{I_n\}} f'(x)dx + \sum_{n=1}^{\infty} \int_{I_n} f_1'(x)dx. \tag{17}$$

Suppose the extremities of the subinterval $I_n$ to be $\alpha_n$, $\beta_n$. We may then write

$$\int_{I_n} f_1'(x)dx = \int_{\alpha_n}^{\beta_n} f_1'(x)dx = f_1(\beta_n) - f_1(\alpha_n) = f(\beta_n) - f(\alpha_n).$$

It is to be noted that $\sum_{n=1}^{\infty} \{f(\beta_n) - f(\alpha_n)\}$

is nothing else than the sum of the variation of $f(x)$ in the intervals of the set $\{I_n\}$. Because the function $f(x)$ is absolutely continuous, it follows that this sum approaches zero with the measure of the set $\{I_n\}$; that is, with the sum of the lengths of these intervals. But as the sum of the lengths of the intervals $\{I_n\}$ approaches zero, the complementary set $C\{I_n\}$ becomes the interval $(a, b)$ except for a set of measure zero. Since $f'(x)$ is summable in the interval $(a, b)$ and the value of the integral is not affected by the omission of a set of points of measure zero, we have from (17)

$$f(b) - f(a) = \int_a^b f'(x)dx,$$

which is the desired result.

This theorem shows that a necessary and sufficient condition that a function $f(x)$ may be the indefinite integral of a summable function is that $f(x)$ shall be absolutely continuous in the interval of integration. The theorem holds if the derivative $f'(x)$ is replaced by any derivative number, say $D^+f(x)$.

For bounded functions, we have seen that whenever the Riemann integral exists, the Lebesgue integral must also exist, and the two are equal. By aid of the foregoing theorem we are able to show that this is not necessarily true for non-bounded functions; that is, there are non-bounded functions which can be integrated in the Riemann sense, but which do not possess an integral in the Lebesgue sense.

For example, consider the function obtained by differentiating the function

$$f(x) = x^2 \sin \frac{1}{x^2}, \quad \text{for } x \neq 0$$
$$= 0, \quad \text{for } x = 0;$$

that is, the function

$$f'(x) = 2\left[x \sin \frac{1}{x^2} - \frac{1}{x} \cos \frac{1}{x^2}\right], \quad x \neq 0, \text{ where } f'(0) = 0.$$

As we have seen, $f'(x)$ is a measurable function,[1] although it is not bounded in any interval including the origin. Moreover, $f(x)$ is not absolutely continuous, since in every neighborhood of the origin the sum of the oscillations exceeds every finite bound,[2] however small a neighborhood may be chosen. Consequently, it follows from the foregoing theorem that $f'(x)$ is not summable. It is, however, integrable in the Riemann sense.

As we have seen, all bounded, measurable functions are summable. The foregoing illustration shows that this is not always the case for non-bounded functions; for, as pointed out, the function $f(x)$ is measurable, although non-bounded and non-summable.

One wishes to know the conditions under which the fundamental theorem of integral calculus holds for the Lebesgue integral; that is, the conditions under which a knowledge of the derivative (or a derivative number) of a function determines that function except as to an additive constant. From the foregoing discussions we are now able to say that a continuous function is determined except as to an additive constant if its derivative (or one of its derivative numbers) is bounded for all values of $x$ in the given interval $(a, b)$. As the foregoing illustration shows, this relation, however, no longer necessarily holds if the derivative (or derivative number) is not bounded in $(a, b)$. We can also say that the fundamental theorem of the integral calculus holds if we know the given function to be absolutely continuous, without placing any condition upon the derivative (or derivative number) other than its existence.

It follows also as a consequence of Theorem I that every indefinite Lebesgue integral defines a continuous function of limited variation. The converse of this statement, however, is not true, although it is true that every continuous function of limited variation has a summable derivative.[3]

---

[1] See Art. 49.

[2] See Art. 38.

[3] See foot-note on p. 320.

For the case of Lebesgue integrals integration may be regarded as the inverse of differentiation for functions having a bounded derivative (or derivative number), but this relation does not necessarily hold if the derivative (derivative number) is not bounded, even though it may be finite in the given interval.

**55. Lebesgue integrals of functions of two independent variables.** In Art. 50, we have seen how the Lebesgue double integral may be defined for a two-dimensional point set $E$. In case $f(x, y)$ is measurable and bounded on the given set, it follows as for functions of a single variable that the Lebesgue double integral exists. The proof is similar to that for a function of a single variable.

The notion of the Lebesgue double integrals may be extended to non-bounded functions of two or more variables in precisely the same manner as was done in Art. 53 for the case of integrals of a single variable. A very large number of the conclusions with reference to the integrals of a single variable may be generalized to include the integrals of functions of two variables, as the character of the proofs does not depend upon the number of variables involved.

We shall consider the conditions under which a double Lebesgue integral may be made to depend upon iterated single integrals. To do this we shall make use of the **Borel notion of measure.** A set is said to be measurable ($B$) if it is formed from closed intervals or rectangles by addition and taking complements, repeated a finite or if necessary an enumerably infinite number of times. Consequently, the sum, difference, and product of sets which are measurable ($B$) are themselves measurable ($B$). Moreover, the limit of a sequence of sets measurable ($B$) is also measurable ($B$) by virtue of this definition. It follows also that if a two-dimensional set is measurable ($B$), then on any straight line in the given domain the linear subset of $E$ on this line is also measurable ($B$), a conclusion which is not always true for two-dimensional sets measurable in the Lebesgue sense. A function $f(x, y)$ is measurable ($B$) in the region $R$ if for all finite values of $A$ the sets

$$E(f \geqq A),\ E(f < A),\ E(f > A),\ E(f \leqq A)$$

are measurable ($B$). We can also say that $f(x, y)$ is measurable ($B$) on a set $E$ if it is measurable ($B$) in a region $R$ provided $f(x, y)$ takes the value zero for all points of this region belonging to the complementary set $C(E)$.

In the discussion to follow, we shall make use of what is called the **characteristic function of a set.** For example, $\phi(x, y)$ is called a characteristic function of the two-dimensional set $E$ if it takes the value one at all points of $E$ and the value zero at the complementary set $C(E)$. Such a function obeys certain laws. Thus if $\phi$ is the characteristic function of the set $E$, then $1 - \phi$ is the characteristic function of $C(E)$. Moreover, if $\phi_1, \phi_2, \ldots$ are the characteristic functions respectively of the sets $E_1, E_2, \cdots$, no two of which have points in common, then

$$\phi = \phi_1 + \phi_2 + \cdots$$

is the characteristic function of the set $E$, where

$$E = E_1 + E_2 + \cdots;$$

and

$$\phi = \phi_1 \cdot \phi_2 \cdot \cdots$$

is the characteristic function of the product

$$E = E_1 \cdot E_2 \cdot \cdots.$$

We have also the following theorem.

THEOREM I. *A characteristic function $\phi$ of a set $E$, which is measurable $(B)$, is a Baire function.*

By a Baire function is understood one which belongs to one of the Baire classes. As already pointed out, a set $E$ is measurable $(B)$ if it arises from the combination of closed intervals and rectangles or by taking the complements of such domains, including the passing to a limit if necessary. The characteristic function of the sets thus obtained by combining in a similar manner the characteristic functions of the individual sets gives the characteristic function of the resulting set. To establish the theorem, it is therefore sufficient to show that the characteristic function of any closed region is a Baire function.

To do this, denote by $d$ the distance of any point $P$ from the given region $R$. Then $d$ is a continuous function of $P$ and takes the value zero whenever $P$ lies in $R$. We may then define the characteristic function of $R$ by the relation

$$\phi = \underset{n \to \infty}{L} \phi_n, \text{ where } \phi_n \equiv \frac{1}{1 + nd}.$$

As $\dfrac{1}{1 + nd}$ is continuous in $d$, $\phi$ is a Baire function of class one, since it is the limit of a sequence of continuous functions and is not itself a continuous function.

We shall now prove the following theorem.

THEOREM II. *Any function which is measurable* ($B$) *is a Baire function.*

As may be readily seen, the limiting function of a sequence of functions each of which is measurable ($B$) is itself measurable ($B$). It is sufficient, therefore, to demonstrate the theorem for bounded functions, since every non-bounded function may be regarded as the limit as $n$ becomes infinite of a sequence of functions each of which is the same function bounded by the condition that

$$|f(x, y)| < n.$$

Let $f(x, y)$ be bounded and measurable ($B$). Divide the interval between the least upper and the greatest lower bounds of $f(x, y)$ into $n$ intervals by taking on the $Z$-axis the points

$$c_1, c_2, \cdot \cdot \cdot, c_k, c_{k+1}, \cdot \cdot \cdot, c_{n+1},$$

where they are so taken that $c_{k+1} - c_k$ is less than $\epsilon$, $c_1$ being the greatest lower bound and $c_{n+1}$ the least upper bound of $f(x, y)$. Denote by $\phi_k$ the characteristic function of the set $E[c_k \leqq f(x, y) < c_{k+1}]$. By Theorem I each function $c_k\phi_k$ is a Baire function. Moreover, the given function $f(x, y)$, which is the limit of the sequence of functions of the type $\{F_n(x, y)\}$, where

$$F_n(x, y) \equiv c_1\phi_1 + c_2\phi_2 + \cdot \cdot \cdot + c_k\phi_k + \cdot \cdot \cdot + c_n\phi_n,$$

must also be a Baire function, as the theorem requires.

We can now establish the following theorem.

THEOREM III. *Given the region $R$, defined by the inequalities*

$$a \leqq x \leqq b, \ c \leqq y \leqq d.$$

*If $f(x, y)$ is bounded and measurable* ($B$) *in $R$, the double Lebesgue integral taken over $R$ is equal to the iterated integrals taken over the same region; that is,*

$$\int\int_R f(x, y)d(x, y) = \int_a^b dx \int_c^d f(x, y)dy. \qquad (1)$$

The integrand $f(x, y)$ is bounded by hypothesis. If it is also continuous in $(x, y)$ together, then the theorem holds; for, in this case the Riemann double integral exists and the order of integration may be interchanged. Consequently, the same is true for the Lebesgue integral.

Let us now consider the more general case in which $f(x, y)$ is bounded and measurable ($B$). By Theorem II, any function which is measurable $(B)$ belongs to one of Baire's classes of functions. It is sufficient, therefore, to prove the theorem for the case in which $f(x, y)$ is bounded and is one of Baire's functions, say of class $\alpha$. For this purpose we assume that $f(x, y)$ is the limit of a sequence of functions $\{f_n(x, y)\}$ each of which is of class less than $\alpha$ but bounded simultaneously as to $n$ and $(x, y)$ and measurable $(B)$. As $f_n(x, y)$ is measurable $(B)$ in $R$, it is also measurable $(B)$ on an interval $(c, d)$ on any line lying in $R$ and parallel to the $Y$-axis. By the extension of Theorem VII of Art. 51 to functions of two variables, we have

$$\int\int_R f(x, y)d(x, y) = \underset{n\to\infty}{L} \int\int_R f_n(x, y)d(x, y),$$

$$\int_c^d f(x, y)dy = \underset{n\to\infty}{L} \int_c^d f_n(x, y)dy.$$

If we assume that the integrals

$$\int\int_R f_n(x, y)d(x, y), \int_c^d f_n(x, y)dy$$

exist for the functions $f_n(x, y)$ of class less than $\alpha$ and that the relation (1) holds for each $f_n(x, y)$ we have from the foregoing relations

$$\begin{aligned}\int\int_R f(x, y)d(x, y) &= \int_a^b dx \underset{n\to\infty}{L} \int_c^d f_n(x, y)dy \\ &= \int_a^b dx \int_c^d f(x, y)dy,\end{aligned}$$

which holds where $f(x, y)$ is of class $\alpha$, as the theorem requires.

Even if $f(x, y)$ is non-bounded, it may be shown that the relation given in (1) holds, provided $f(x, y)$ is non-negative and measurable $(B)$. If we assume that $f(x, y)$ is summable in the domain $R$, then the given theorem can be extended to non-bounded functions having either sign, since the given function can be written as the difference of two non-negative, summable functions.[1]

In case the given function is measurable in the Lebesgue sense, but is not measurable $(B)$, one may still show that the iterated integrals

$$\int_a^b dx \int_c^d f(x, y)dy, \quad \int_c^d dy \int_a^b f(x, y)dx$$

both exist and have the value of the double Lebesgue integral taken over the domain $R$, provided this double integral exists. In reaching this conclusion one makes use of the fact that even though either of the integrals

---

[1] See Lebesgue, *Annali di Mat.*, (3) Vol. VII (1902), pp. 274–282; also de la Vallée Poussin, *Cour d'analyse infinitesimale*, Vol. 2, 2d ed., pp. 117–122.

$$\int_a^b f(x, y)dx, \quad \int_c^d f(x, y)dy$$

taken with respect to the one variable fails to exist for a set of values of measure zero of the remaining variable, the iterated integrals may still be said to exist.[1] It is to be noted, however, that the converse of this statement is not true; that is, there are functions $f(x, y)$ for which the two iterated integrals exist but for which the double Lebesgue integral over the given region does not exist. The extension of the foregoing conclusion to non-bounded functions having either positive or negative values follows from the same considerations as in the previous case.

**56. Other definitions of integration.** Aside from the definitions of integration already considered, there are various others to be found in the mathematical literature. Some of these definitions are generalizations of earlier definitions, while others introduce new conceptions of integration.[2]

Stieltjes and Hellinger have generalized the notion of integration in that the variable with respect to which the integration is taken is replaced by a function.

Thus **Stieltjes** considers two functions $f(x)$ and $\phi(x)$ both of which are bounded. He then divides the given interval $(a, b)$ into $n$ parts by inserting, as in the case of the Riemann integral, the intermediate points

$$x_1, x_2, x_3, \cdot \cdot \cdot, x_{n-1},$$

and sets up the sum

$$\sum_{k=1}^{n} f(\xi_k)\,[\phi(x_k) - \phi(x_{k-1})],$$

where $\xi_k$ is any point interior to the interval $(x_{k-1}, x_k)$. If this sum approaches a definite limit as the maximum value of $|x_k - x_{k-1}|$ approaches zero, the Stieltjes integral is said to exist and is denoted by

$$\int_a^b f(x)\,d\phi(x).$$

If $f(x)$ is continuous and $\phi(x)$ is of limited variation the Stieltjes integral of $f(x)$ with respect to $\phi(x)$ can be shown to exist. It may be shown that every Stieltjes integral can be transformed into a

---

[1] See Fubini, *Rend. del. R. Acc. d. Lincei,* (5) Vol. 16 (1907), pp. 608–614.

[2] For a general discussion of the integrals related to the Lebesgue integral, see, Hildebrandt, *Bull. Amer. Math. Soc.*, Vol. 24 (1917–18), pp. 113–144; pp. 177–202.

Lebesgue integral of another function and conversely.[1] By means of this relation to the Lebesgue integral, the Stieltjes integral may be extended to the case where $f(x)$ is a discontinuous function.[2]

**Hellinger's** definition of an integral is closely related to that of Stieltjes. He considers two functions $f(x)$ and $\phi(x)$, where $f(x)$ is continuous in $(a, b)$ and $\phi(x)$ is continuous and monotone increasing. Moreover, it is assumed that in any portion of $(a, b)$ where $\phi(x)$ is constant the function $f(x)$ is also constant, and consequently if $[\phi(x_2) - \phi(x_1)]$ is zero then $[f(x_2) - f(x_1)]$ is also zero. We now divide the interval $(a, b)$ into a finite number of intervals by means of the intermediate points

$$x_1, \ x_2, \ \cdot \ \cdot \ \cdot, \ x_{n-1},$$

and form the sum

$$\sum_{k=1}^{n} \frac{[f(x_k) - f(x_{k-1})]^2}{\phi(x_k) - \phi(x_{k-1})},$$

the quotient here expressed being defined to be zero whenever the denominator is zero. The least upper bound of this sum for all possible methods of subdivision of the given interval is called the Hellinger integral and is denoted by

$$\int_a^b \frac{[df(x)]^2}{d\phi(x)}.$$

It has been shown that the Hellinger integral may be expressed in terms of a Lebesgue integral, and conversely every Lebesgue integral can be expressed in terms of a Hellinger integral.[3]

**Denjoy** has given an important generalization of the Lebesgue integral.[4] One can best indicate the nature of this integral by first stating certain principles of construction which the author formulates. For the general form of the definition these principles may be stated as follows.

---

[1] See Lebesgue, *Comptes rendus*, Vol. 150 (1910), pp. 86–88; also Van Vleck, *Trans. Amer. Math. Soc.*, Vol. 18 (1917), pp. 326–330; Bliss, *Bull. Amer. Math. Soc.*, Vol. 24 (1917), pp. 28–29.

[2] See Lebesgue, *Comptes rendus*, Vol. 150 (1910), pp. 86–88; also W. H. Young, *Proc. London Math. Soc.* (2) Vol. 13 (1914), pp. 109–150; Radon, *Sitzungsberichte der Akademie der Wissenshaften, Wein*, IIa, Vol. 121 (1912), pp. 265–297.

[3] See Hahn, *Monatshefte für Math. und Phys.*, Vol. 23 (1912), pp. 170–183. For an extension of the Hellinger integral, see Radon, *Sitzungsberichte der Akademie der Wissenschaften, Wein*, II$_a$, Vol. 122 (1913); pp. 1322–32.

[4] See *Comptes rendus*, Vol. 154 (1914), pp. 859–62, and pp. 1075–78. See also Khintchine, *Comptes rendus*, Vol. 162 (1916), pp. 287–91; Denjoy, pp. 377–80, and *Ann. Ecole Norm.* (3), Vol. 34 (1917), pp. 181–238. Denjoy used the term *totalization* in connection with his theory of integration much as the term *summability* is used in connection with the Lebesgue integral.

1. In any subinterval $(\alpha, \beta)$ of $(a, b)$ in which the given measurable function $f(x)$ is summable in the Lebesgue sense the Denjoy integral shall be identical with that of Lebesgue. The same shall also be true on any perfect set.

To distinguish the Denjoy integral from the Lebesgue integral in this discussion we shall denote them respectively by the symbols

$$D\int_a^b f(x)dx, \quad L\int_a^b f(x)dx.$$

2. If the integral $D\int_{\alpha'}^{\beta'} f(x)dx$ is known for all values of $\alpha' < \beta'$ contained in $(\alpha, \beta)$, then we shall have

$$D\int_\alpha^\beta f(x)dx = \underset{\substack{\alpha'\to\alpha\\ \beta'\to\beta}}{L}\ D\int_{\alpha'}^{\beta'} f(x)dx.$$

3. If the Denjoy integral is known for a finite number of consecutive intervals $(\alpha_1, \alpha_2)(\alpha_2, \alpha_3) \cdot \cdot \cdot (\alpha_{n-1}, \alpha_n)$, then we shall have the relation

$$D\int_{\alpha_1}^{\alpha_n} f(x)dx = D\int_{\alpha_1}^{\alpha_2} f(x)dx + D\int_{\alpha_2}^{\alpha_3} f(x)dx + \cdot \cdot \cdot + D\int_{\alpha_{n-1}}^{\alpha_n} f(x)dx.$$

4. Let $E$ be a perfect set of points contained in a subinterval $(\alpha, \beta)$ of $(a, b)$, and suppose that $f(x)$ is summable on the set $E$. Let it be supposed that the Denjoy integral has been defined on every subinterval $(\alpha', \beta')$ of $(\alpha, \beta)$ which contains no points of $E$ as inner points. Let $(\alpha_n, \beta_n)$, $n=1, 2, 3 \cdot \cdot \cdot$ be the set of intervals complementary to $E$ with respect to $(\alpha, \beta)$. Denote by $M_n$ the upper limit of $\left| D\int_{\alpha'}^{\beta'} f(x)dx \right|$ for all intervals $(\alpha', \beta')$ in $(\alpha_n, \beta_n)$. It is assumed that the series $\Sigma M_n$ converges. The Denjoy integral for $(\alpha, \beta)$ is then defined by the relation

$$D\int_\alpha^\beta f(x)dx = \sum_{n=1}^{\infty} D\int_{\alpha_n}^{\beta_n} f(x)dx + L\int_E f(x)dx.$$

The Denjoy integral may then be said to exist for the interval $(a, b)$ by the repeated application of the foregoing principles of construction, provided the given function $f(x)$ satisfies the following conditions.

I. If $E$ is any perfect set contained in $(a, b)$, the points of $E$ in whose neighborhood $f(x)$ is not summable shall form a subset which in no portion of $(a, b)$ is everywhere dense with respect to $E$. A function is not summable in the neighborhood of a point, if there exists no subinterval containing the point on which the function is summable.

II. If $D\int_{\alpha'}^{\beta'} f(x)dx$, $\alpha' < \beta'$ is known for all values of $\alpha'$, $\beta'$ in $(\alpha, \beta)$, then $f(x)$ must be such that the limit $\underset{\substack{\alpha'\to\alpha \\ \beta'\to\beta}}{L}\, D\int_{\alpha'}^{\beta'} f(x)dx$ shall exist.

III. If $E$ is a perfect set in no subinterval everywhere dense and if $D\int_{\alpha_n}^{\beta_n} f(x)dx$ is known in those subintervals $(\alpha_n, \beta_n)$ which are free from points of $E$, then $f(x)$ shall be such that the points of $E$ in the neighborhood of which the series $\sum\limits_{n=1}^{\infty} D\int_{\alpha_n}^{\beta_n} f(x)dx$ does not converge absolutely shall not be everywhere dense with respect to $E$ in any subinterval.

If $f(x)$ satisfies these conditions, then by aid of the foregoing principles of construction the Denjoy integral can be calculated by means of an enumerably infinite set of Lebesgue integrals and passage to the limit.

The Denjoy integral is more general than the Lebesgue integral, in the sense that there exist functions which have no Lebesgue integral which can be integrated by the method of Denjoy. Such an integral is, however, not absolutely convergent; that is, the Denjoy integral of the absolute value of the function does not exist. Moreover, for these integrals, we can say, as we did in case of Lebesgue integrals, that the indefinite integral of $f(x)$ is continuous and has $f(x)$ as a derivative with the exception of at most a set of points of measure zero.

**Perron**[1] has introduced a notion of integration quite different from all of the foregoing definitions in that he has not based it on a summation process but has considered integration as the inverse operation of differentiation. By introducing what he designates as the adjoined upper and lower functions he is able to formulate a definition of integration which for bounded functions is identical with the Lebesgue integral, but for non-bounded functions leads to a more general class of integrable functions than is the case with the Lebesgue integral. Under certain restrictions as to the continuity of the adjoined upper and lower functions, the Perron integral for non-bounded functions becomes identical with that of Lebesgue. The Perron integral can be readily extended to functions of two or more variables.

---

[1] See *Sitzungsberichte der Heidelberg Akademie der Math. Wiss.* (1914), Vol. 14, pp. 1–16. For other citations concerning the literature on the Perron integral, see *Encyklop. der Math. Wiss.* Vol. $\text{II}_3$, Part 7, pp. 1074–76.

Borel[1] has given a generalized definition of integration which for bounded functions is more restricted in its applications than the Lebesgue integral, but for non-bounded functions can be applied to non-absolutely convergent integrals, which is not the case with the Lebesgue integrals. When, however, both the Borel and the Lebesgue integrals exist, they give the same value.

Other modifications of the definitions of integration by Riemann and Lebesgue have been proposed by W. H. Young[2] and by Pierpont.[3] For example, if we disregard the case where the upper and the lower integrals may become infinite, it follows that the Young integral leads to the same class of integrable functions as the Lebesgue definition of integration. The Pierpont definition differs from that of Lebesgue in that the set of points over which the integral is taken is not necessarily a measurable set.

## EXERCISES

**1.** Is the function $f(x) = \dfrac{1}{1 - \dfrac{1}{x}}$ a measurable function in the interval $(1, -1)$?

**2.** State the fundamental theorem of integral calculus. Compare the conditions under which it holds for Riemann integrals and for Lebesgue integrals.

**3.** Given a function of limited variation. Is such a function integrable in the Riemann sense? in the Lebesgue sense? Could it have a finite derivative at the rational points and not at the irrational points? at the irrational points and not at the rational points?

**4.** In the Lebesque theory of integration would any rational integral function of integrable functions be itself an integrable function?

**5.** Given $f(x) = \underset{n\to\infty}{L} \; f_n(x)$, where $f_n(x)$ is a point-wise discontinuous function. Under what conditions is $f(x)$ summable?

**6.** Show that every upper semi-continuous function is measurable and integrable. What statement can be made concerning the saltus function $\omega[f(x)]$?

---

[1] *Comptes rendus*, Vol. 150 (1910), pp. 375–377; also *Encyklop. der Math. Wiss.*, Vol. $\text{II}_3$, Part 7, p. 1064.

[2] *Philos. Trans.* Vol. 204A (1905), pp. 221–252; *Proceedings Royal Philos. Soc.*, Vol. 88A (1913), pp. 170–178, *Proceedings London Math. Soc.*, Series 2, Vol. 13 (1914), pp. 109–150.

[3] *Theory of Functions of Real Variables*, Vol. II (1912), pp. 343 *et seq.*

**7.** Is every function which is continuous with respect to a set of points everywhere dense an integrable function in the Lebesgue sense? Is such a function necessarily measurable?

**8.** Show that the function $\phi(x) = 2x \sin \frac{1}{x} - \frac{2}{x} \cos \frac{1}{x}$ is not summable. [See Art. 49; also Lebesgue, *Leçons sur l'intégration*, p. 115.]

**9.** In Ex. 4, Art. 25, we considered the function $f(x) = \underset{n\to\infty}{L} \; \underset{t\to\infty}{L} \frac{\sin^2 n!\pi x}{\sin^2 n!\pi x + t^2}$. Is this function bounded? measurable? integrable in the Riemann sense? in the Lebesgue sense?

**10.** Let $f(x)$ be defined as the limit of a sequence $\{f_n(x)\}$, where $f_n(x)$ has only ordinary discontinuities. Is $f(x)$ integrable in the Lebesgue sense?

**11.** Let $f(x)$ be any Baire function and define $\phi(x)$ by the relation $\phi(x) = \int_a^x f(x)dx$. Show that $\frac{d\phi}{dx} = f(x)$ at all points, with the exception of at most a set of measure zero.

**12.** If $f(x)$ is summable, show that the indefinite integral $\phi(x) = \int_a^x f(x)dx$ can be expressed as the difference of two monotone functions.

**13.** If $f(x)$ is summable over the interval $(a, b)$, show that it is summable over the set $E$ consisting of those points of $(a, b)$ which correspond to the algebraic numbers. Show that this integral is zero.

**14.** Show that Theorem III, Art. 40, can be extended to Lebesgue integrals.

**15.** Show that if $f(x)$ is integrable in the Lebesgue sense on the set $E$, it is integrable on any measurable subset of $E$.

**16.** Define $f(x)$ as equal to zero for all rational values of $x$ and equal to $\frac{1}{x}$ for all irrational values. Is $f(x)$ measurable? Is it summable?

**17.** Given a function $f(x, y)$ which is continuous in each variable. Does the double Lebesgue integral exist? Do the two iterated integrals exist?

**18.** Is every rational function measurable? Is it bounded? Is it integrable in the Lebesgue sense?

**19.** Given the function $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ for $x \neq 0, y \neq 0$, where $f(x, y) = 0$, for $x = 0, y = 0$. Does the double integral $\int_0^1 \int_0^1 f(x, y)\, d(x, y)$ exist? Do the iterated integrals $\int_0^1 dx \int_0^1 f(x, y)dy, \int_0^1 dy \int_0^1 f(x, y)dx$ exist, and if so, are they equal?

# CHAPTER VII

## INFINITE SERIES

**57. General definitions.** Suppose we have given an infinite succession of positive or negative constants

$$u_1, u_2, u_3, \cdot \cdot \cdot, u_n, \cdot \cdot \cdot, \qquad (a)$$

whose order of succession follows some general law. The expression

$$\sum_{n=1}^{\infty} u_n \equiv u_1 + u_2 + \cdot \cdot \cdot + u_n + \cdot \cdot \cdot,$$

where, as indicated, the summation is extended indefinitely, is called an infinite series, or more briefly a **series**. The numbers

$$u_1, u_2, u_3, \cdot \cdot \cdot, u_n, \cdot \cdot \cdot$$

are called the **terms** of the series. It is convenient to put

$$S_n = u_1 + u_2 + \cdot \cdot \cdot + u_n.$$

The series is said to **converge** if $S_n$ has a definite limit as $n$ becomes infinite, that is if

$$\underset{n\to\infty}{L} S_n = A,$$

and $A$ is called the **sum** of the series $\sum\limits_{n=1}^{\infty} u_n$, often written $\Sigma u_n$. If the limit does not exist, then the series is said to be **divergent.**

A necessary condition for the convergence of a series $\sum\limits_{n=1}^{\infty} u_n$ is that

$$\underset{n\to\infty}{L} u_n = 0;$$

for, otherwise the limit of $S_n$ could not be said to exist. This condition is not sufficient, however; for, the harmonic series

$$\Sigma \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdot \cdot \cdot + \frac{1}{n} + \cdot \cdot \cdot$$

can be easily shown to be divergent, although it satisfies the above condition.

We should distinguish between a series and a sequence. An infinite sequence is merely a succession of terms following some general law in the order of succession. An infinite series is the limit of the sum of such a succession of terms. The two are related in that the limit or sum of a series is the limit of the sequence

$$S_1,\ S_2,\ S_3,\ \cdot\ \cdot\ \cdot,\ S_n,\ \cdot\ \cdot\ \cdot$$

as $n$ becomes infinite. Hence, the laws of limits of sequences may be made to apply to the limits of series.

Instead of the terms of the series being constants, they may be functions of one or more variables. Thus, if $u_n$ is a function of $x$, we may have the series

$$\Sigma\, u_n(x) = u_1(x) + u_2(x) + \cdot\ \cdot\ \cdot + u_n(x) + \cdot\ \cdot\ \cdot\ .$$

Such a series is said to converge for $x = x_0$, if after having replaced $x$ by $x_0$, the series of constant terms $\Sigma\, u_n(x_0)$ is convergent. If the series converges for each value of $x$ in an interval $(a, b)$, we say that $\Sigma\, u_n(x)$ converges in this interval, which is called the **interval of convergence** if the given series converges for all values of $x$ within the interval and diverges for all values exterior to it. The interval of convergence may be closed or not; that is, the question of convergence at the extreme points of the interval depends upon the nature of the given series.

In the interval of convergence $(a, b)$, the series may be said to define a function, and we may write

$$f(x) = \underset{n\to\infty}{L}\ S_n(x) = \sum_1^{\infty} u_n(x).$$

The function $f(x)$ is then defined for all values of $x$ in the interval $(a, b)$, which may be either open or closed. Thus, the series

$$1 + x + x^2 + \cdot\ \cdot\ \cdot + x^n + \cdot\ \cdot\ \cdot$$

converges for all values of $x$ within the interval $(-1, 1)$ and we have

$$f(x) = \underset{n\to\infty}{L}\ S_n(x) = \underset{n\to\infty}{L}\ \frac{1 - x^n}{1 - x} = \frac{1}{1 - x}.$$

In this case, the series defines $f(x)$ only for the open interval since the series does not converge for $|\,x\,| \geqq 1$.

If the terms of the series are functions of two variables, the convergent series defines likewise a function of two variables at each point of the region of convergence.

In order to give a geometric interpretation to the convergence of a series, let the coördinates of a point in space be $(x, \frac{1}{n}, y)$. If the series

$$u_1(x) + u_2(x) + \cdot \cdot \cdot + u_n(x) + \cdot \cdot \cdot ,$$

converges for $x = x_0$, or what is the same thing, if the limit $\underset{n\to\infty}{L} S_n(x_0)$ exists, then for a given $\epsilon$ there must exist

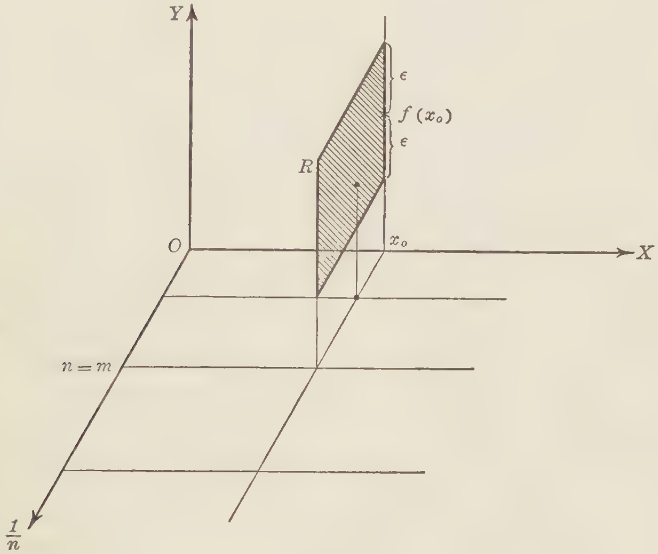


FIG. 41.

in the plane through $x_0$ parallel to the $\left(Y, \frac{1}{n}\right)$-plane a rectangle $R$, as indicated in Fig. 41, such that for all values of $n > m$, the corresponding values of $S_n(x_0)$ lie within this rectangle; that is, we have

$$| f(x_0) - S_n(x_0) | < \epsilon, \quad n > m.$$

The sides of this rectangle are given by the equations

$$n = \infty, \ n = m, \ y = f(x_0) \pm \epsilon.$$

The terms of a series may not all be positive. The series formed by replacing each term by its absolute value is sometimes called the **adjoint series.** If the adjoint series converges, then the given series is said to **converge absolutely.** Thus the series

$$1 - \frac{1}{2} + \frac{1}{2^2} - \frac{1}{2^3} + \cdot \cdot \cdot ,$$

converges absolutely, since the series formed by taking each term positively, namely

$$1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdot \cdot \cdot ,$$

converges.

If the series contains both positive and negative terms and converges, but not absolutely, it is said to be **conditionally convergent.** The following is such a series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdot \cdot \cdot .$$

A conditionally convergent series is characterized by the fact that the positive terms taken by themselves do not converge and likewise the negative terms taken by themselves do not form a convergent series. For, if the series of positive terms and the series of negative terms should both converge, then the given series must, by the law of limits, converge absolutely and to the algebraic sum of the limits of the two series. If either the series of positive terms or the series of negative terms fails to converge while the other converges, it follows that the given series can not converge.

On the other hand, both the positive terms and the negative terms of a conditionally convergent series may form by themselves divergent series. In this case, it is possible, by a proper rearrangement of the terms, to make the given series converge to any arbitrarily chosen number, or even to diverge. For example, consider a conditionally convergent series $\Sigma\, u_n$; and let $\Sigma\, a_n$ be the series of its positive terms and $\Sigma\, b_n$ the series of its negative terms. Since the given series $\Sigma\, u_n$ converges, it follows that the terms of the series $\Sigma\, a_n$, $\Sigma\, b_n$ must have the property that

$$\underset{n \to \infty}{L}\ a_n = 0, \quad \underset{n \to \infty}{L}\ b_n = 0.$$

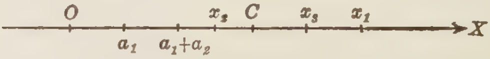


FIG. 42

Let $C$ be any arbitrarily assigned number. It may be denoted by a point on a straight line. Denote by points on the line $OX$ the numbers of the sequence

$$a_1,\ a_1 + a_2,\ a_1 + a_2 + a_3,\ \cdot \cdot \cdot . \tag{1}$$

Since $\Sigma\, a_n$ is a divergent series some of these points must lie on the opposite side of $C$ from the origin $O$. Let $x_1$ be the first such point, and let the corresponding term of the sequence (1) be

$$a_1 + a_2 + a_3 + \cdot \cdot \cdot + a_k.$$

To this result add a sufficient number of terms from the series of negative terms, $\Sigma\, b_n$, so that we ultimately get a point $x_2$ lying between $O$ and $C$. We then add terms from $\Sigma\, a_n$ $(n > k)$ until we get a point lying again on the opposite side of $C$ from $O$. Let $x_3$ be the first such point obtained. We then add more terms from $\Sigma\, b_n$ and get again a point lying on the left of $C$. Continuing in this way we get a set of points dense at $C$ because of the condition that the terms $a_n$, $b_n$ approach zero as $n$ becomes infinite. Consequently, the series as thus arranged converges to the required value.

Since we can always so arrange the terms of the foregoing series $\Sigma\, u_n$ in such a manner that the resulting series exceeds any given number $A$, however large this number may be chosen, it follows that by a proper arrangement of the terms of a conditionally convergent series $\Sigma\, u_n$ we can form a divergent series.

In the use of infinite series, it must not be assumed that the properties which hold for the sum of a finite number of terms may be extended without modification to an infinite series. Thus, the sum of a finite number of continuous functions is always a continuous function. The corresponding conclusion is not always true if the number of terms becomes infinite. Likewise, we can differentiate or integrate term by term the sum of a finite number of terms and thus obtain the derivative or the integral of the function represented by the sum, but it does not follow that we can do the same thing with an infinite series of terms even though the individual terms may be differentiated or integrated. In this chapter, we shall be interested particularly in discussing the conditions under which the processes of the calculus can be extended to infinite series. Many of these questions depend in the last analysis upon the interchange of the order of limits, that is upon whether the limit of the sum is the sum of the limits. The following illustration shows that this is not always the case.

Consider the series where

$$S_n = \left\{\left(\frac{1}{2} + \frac{k+1}{k^2}\right) + \left(\frac{1}{2^2} + \frac{k+2}{k^2}\right) + \cdot\ \cdot\ \cdot + \left(\frac{1}{2^n} + \frac{k+n}{k^2}\right)\right\}.$$

We have as the sum of the limits of the terms of the series

$$\underset{n\to\infty}{L}\ \underset{k\to\infty}{L}\ S_n = \underset{n\to\infty}{L}\left\{\underset{k\to\infty}{L}\left(\frac{1}{2} + \frac{k+1}{k^2}\right) + \underset{k\to\infty}{L}\left(\frac{1}{2^2} + \frac{k+2}{k^2}\right) + \cdot\ \cdot\ \cdot \right.$$
$$\left. + \underset{k\to\infty}{L}\left(\frac{1}{2^n} + \frac{k+n}{k^2}\right)\right\}$$
$$= \frac{1}{2} + \frac{1}{2^2} + \cdot\ \cdot\ \cdot + \frac{1}{2^n} + \cdot\ \cdot\ \cdot = 1.$$

On the other hand, the limit $\underset{k\to\infty}{L}\ \underset{n\to\infty}{L}\ S_n$ does not exist; for, we have

$$\underset{k\to\infty}{L}\ \underset{n\to\infty}{L}\ S_n = \underset{k\to\infty}{L}\ \underset{n\to\infty}{L}\left\{\frac{1}{2}\,\frac{\dfrac{1}{2^n}-1}{\dfrac{1}{2}-1}+\frac{nk+\dfrac{n(n+1)}{2}}{k^2}\right\}$$

$$= \underset{k\to\infty}{L}\ \{1+\infty\} = \infty.$$

Consequently, the sequential limit taken in this order does not exist.

**58. Laws of operation with series.** As we have seen, the laws of operation with a finite number of terms can not be assumed as holding when the number of terms becomes infinite. Even the processes of arithmetic themselves must be defined anew for this extended domain of operation. It will be seen, however, that with certain modifications and restrictions these processes can be so extended. We shall now consider the conditions under which this may be done.

One of the first questions to arise is as to whether one can always insert parentheses at pleasure in a series without affecting its convergence or changing its sum. This question is answered by the following theorem.

THEOREM I. *The terms of a convergent series can always be united into groups in any manner without affecting the sum of the series, provided the order of the terms is not changed.*

Let $S_n$ be the sum of the first $n$ terms of the given series and denote by $S'_m$ the sum of the first $m$ groups of terms. However large $n$ may be, $m$ may be taken sufficiently large so that $S'_m$ will contain at least all of the terms of $S_n$. If $S'_m$ contains $n + p$ terms of the given series $\Sigma\, u_n$, we have

$$S'_m - S_n = u_{n+1} + u_{n+2} + \cdot\ \cdot\ \cdot + u_{n+p}, \quad p = 1, 2, \cdot\ \cdot\ \cdot\ .$$

Denoting the second member of this equation by ${}_pR_n$, we have for any arbitrarily chosen $\epsilon$

$$|\,{}_pR_n\,| < \epsilon,$$

for all values of $n$ sufficiently large. As $n$ becomes infinite, the value of $m$ also increases without limit. The foregoing inequality is then equivalent to saying that

$$\underset{n\to\infty}{L}\ S_n = \underset{m\to\infty}{L}\ S'_m.$$

This theorem shows that we can always insert parentheses at pleasure in a series. However, it is not always possible to remove them without affecting the sum of the series, as the following example shows.

**Ex. 4.** Given the series

$$\left(\frac{1}{2}-\frac{2}{3}\right)+\left(\frac{2}{3}-\frac{3}{4}\right)+\left(\frac{3}{4}-\frac{4}{5}\right)+\cdot\cdot\cdot+\left(\frac{n}{n+1}-\frac{n+1}{n+2}\right)+\cdot\cdot\cdot$$
$$=-\left\{\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\frac{1}{4\cdot 5}+\cdot\cdot\cdot+\frac{1}{(n+1)(n+2)}+\cdot\cdot\cdot\right\}.$$

This series is convergent, since its terms are less than the corresponding terms of the convergent series $\Sigma\frac{1}{n^2}\cdot$ In fact, the series converges to $-\frac{1}{2}\cdot$ If we remove the parentheses in the given series, we have the oscillating series

$$\frac{1}{2}-\frac{2}{3}+\frac{2}{3}-\frac{3}{4}+\frac{3}{4}-\frac{4}{5}+\cdot\cdot\cdot,$$

which can not have a unique sum, since the limit $\underset{n\to\infty}{L}S_n$ now depends upon whether $n$ is even or odd.

The following theorem gives a sufficient condition for the removal of parentheses.

THEOREM II. *If a series involving parentheses converges absolutely after the parentheses have been removed, then the removal of the parentheses does not affect the sum of the series.*

Denote by $S'_m$ the sum of the first $m$ terms of the given series, where the content of each parenthesis is taken as one term, and let $S_n$ denote the sum of the first $n$ terms after the parentheses have been removed. The value of $m$ may certainly be taken sufficiently large to contain at least the $n$ terms of $S_n$. It follows that

$$|S'_m - S_n| \leqq |u_{n+1}| + |u_{n+2}| + \cdot\cdot\cdot + |u_{n+p}|,$$

where $n+p$ denotes the number of terms of the series which are included in the first $m$ parentheses. From the absolute convergence of the series after the parentheses have been removed, we know that the right-hand member of this inequality may be made as small as we please by taking $n$ sufficiently large, say $n > n'$. We have then for an arbitrarily small $\epsilon$, remembering that as $n$ increases indefinitely $m$ does also,

$$|S'_m - S_n| < \epsilon,\ n > n',\ m > n',$$

whence

$$\underset{m\to\infty}{L} S'_m = \underset{n\to\infty}{L} S_n.$$

THEOREM III. *In an absolutely convergent series, the terms may be rearranged at pleasure without changing the sum of the series.*

First of all, let us assume that all of the terms of the given series are positive. Put

$$S_n = u_1 + u_2 + \cdot \cdot \cdot + u_n,$$

and let

$$\underset{n \to \infty}{L} S_n = A.$$

Denote by $S'_m$ the sum of the first $m$ terms of the rearranged series. As all of the terms are positive $S'_m$ is a monotone increasing function of $m$. Having selected the value of $m$, the value of $n$ can be taken large enough so that all of the terms of $S'_m$ are included in $S_n$. Hence, we have

$$S'_m < S_n < A.$$

Consequently, $S'_m$ is not only a monotone increasing function of $m$, but it is bounded; and it follows that $S'_m$ must have a definite limit as $m$ becomes infinite. We may therefore write

$$\underset{m \to \infty}{L} S'_m = B \leqq A.$$

If we now regard the given series $\Sigma\, u_n$ as a rearrangement of the series $\Sigma\, u'_m$ we get, in a manner similar to that employed above,

$$\underset{n \to \infty}{L} S_n = A \leqq B.$$

Consequently, it follows that $A$ and $B$ must be equal and hence

$$\underset{n \to \infty}{L} S_n = \underset{m \to \infty}{L} S'_m = A.$$

Let us now suppose that the given series contains negative terms but that it converges absolutely. Let $\Sigma\, a_n$ be the series of positive terms and $\Sigma\, b_n$ the series of negative terms. Denote by $S_m(a)$ the sum of the first $m$ positive terms and by $S_p(b)$ the sum of the first $p$ negative terms of the given series. Moreover, let us so select $m$ and $p$ that

$$n = m + p.$$

We have then

$$S_n = S_m(a) - S_p(b).$$

As $m$ and $p$ become infinite, it follows that $n$ also becomes infinite. As the given series converges absolutely, it follows that the two series $\Sigma\, a_n$, $\Sigma\, b_n$ converge. We may then write

$$\underset{m \to \infty}{L} S_m(a) = M, \quad \underset{p \to \infty}{L} S_p(b) = P,$$

whence, noting that

$$\underset{n\to\infty}{L} S_n = \underset{m\to\infty}{L} S_m(a) - \underset{p\to\infty}{L} S_p(b),$$

it follows that $\quad A = M - P.$

Denote the rearranged series by

$$u_1' + u_2' + \cdots + u_m' + \cdots.$$

The series of positive terms and likewise the series of negative terms in $\Sigma u_m'$ must converge by virtue of the foregoing discussion. It follows that $\Sigma u_m'$ must also converge. If we denote the three sums by $M'$, $P'$, $A'$, respectively, we have

$$A' = M' - P'.$$

It still remains to show that $A$ and $A'$ are equal. However, by the first part of this discussion, it follows that

$$M' = M, \ P' = P.$$

Hence, we have $\quad A' = A;$

that is, the theorem is established for the case of absolutely convergent series as well as for series containing only positive terms.

To rearrange the terms of an infinite series is equivalent to extending of the commutative law of addition to the sum of an infinite number of terms. The foregoing theorem gives the condition under which that may be done.

**Ex. 5.** Find the sum of the series

$$\frac{1}{2} + \frac{1}{2^3} - \frac{1}{2^2} + \frac{1}{2^5} + \frac{1}{2^7} - \frac{1}{2^4} + \frac{1}{2^9} + \frac{1}{2^{11}} - \frac{1}{2^6} + \cdots.$$

This series converges absolutely; hence, by the foregoing theorem, the terms can be arranged in any convenient order without affecting the sum. We may then write the series as follows.

$$\begin{aligned}
&\frac{1}{2} - \frac{1}{2^2} + \frac{1}{2^3} - \frac{1}{2^4} + \frac{1}{2^5} - \frac{1}{2^6} + \frac{1}{2^7} - \cdots \\
&= \frac{1}{2}\left\{1 - \frac{1}{2} + \frac{1}{2^2} - \frac{1}{2^3} + \frac{1}{2^4} - \frac{1}{2^5} + \cdots\right\} \\
&= \frac{1}{2}\left\{\left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \cdots\right) - \left(\frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \cdots\right)\right\} \\
&= \frac{1}{2}\left\{\frac{1}{1-\frac{1}{4}} - \frac{1}{2}\cdot\frac{1}{1-\frac{1}{4}}\right\} = \frac{1}{2}\left\{\frac{4}{3} - \frac{2}{3}\right\} = \frac{1}{3}.
\end{aligned}$$

We shall frequently need to combine series and hence we must define such operations as addition, subtraction, multiplication, etc., when applied to infinite series. If we have given two series $\Sigma\, u_n$, $\Sigma\, v_n$ we shall understand by the **sum** of these series that series formed by adding the corresponding terms of the two series: that is,

$$\begin{aligned}\Sigma\, u_n + \Sigma\, v_n &= \Sigma(u_n + v_n)\\ &= (u_1 + v_1) + (u_2 + v_2) + \cdot\cdot\cdot + (u_n + v_n) + \cdot\cdot\cdot .\end{aligned}$$

That the **commutative law of addition** holds for the sum of two series follows at once from the foregoing definition: that is,

$$\Sigma\, u_n + \Sigma\, v_n = \Sigma\, v_n + \Sigma\, u_n.$$

A corresponding definition can be formulated for **subtraction;** namely,

$$\begin{aligned}\Sigma\, u_n - \Sigma\, v_n &= \Sigma(u_n - v_n)\\ &= (u_1 - v_1) + (u_2 - v_2) + \cdot\cdot\cdot + (u_n - v_n) + \cdot\cdot\cdot .\end{aligned}$$

In the multiplication of one polynomial by another, we multiply each term of the one by every term of the other and take the sum of the products formed. For example, if we put

$$\begin{aligned}S_n &= u_1 + u_2 + \cdot\cdot\cdot + u_n,\\ T_n &= v_1 + v_2 + \cdot\cdot\cdot + v_n,\end{aligned}$$

the product of the two polynomials may be written

$$\begin{aligned}S_nT_n = {} & u_1v_1 + u_1v_2 + u_1v_3 + \cdot\cdot\cdot + u_1v_n\\ &+ u_2v_1 + u_2v_2 + u_2v_3 + \cdot\cdot\cdot + u_2v_n\\ &+ u_3v_1 + u_3v_2 + u_3v_3 + \cdot\cdot\cdot + u_3v_n\\ &+ u_4v_1 + u_4v_2 + u_4v_3 + \cdot\cdot\cdot + u_4v_n\\ &\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\\ &+ u_nv_1 + u_nv_2 + u_nv_3 + \cdot\cdot\cdot + u_nv_n.\end{aligned} \tag{3}$$

If $n$ is allowed to become infinite, we are led to the product of two infinite series. It is convenient to exhibit such a product by grouping the terms of the above array by diagonals. We shall understand the **product** of the series $\Sigma\, u_n$, $\Sigma\, v_n$ to mean the series

$$\begin{aligned}\Sigma(u_nv_1 + u_{n-1}v_2 + \cdot\cdot\cdot + u_2v_{n-1} + u_1v_n) = u_1v_1 + (u_2v_1 + u_1v_2)\\ +(u_3v_1+u_2v_2+u_1v_3)+\cdot\cdot\cdot+(u_nv_1+u_{n-1}v_2+\cdot\cdot\cdot+u_1v_n)+\cdot\cdot\cdot .\end{aligned} \tag{4}$$

The question is then under what conditions does this series converge to the product $AB$; that is, under what conditions do we have

$$\underset{n\to\infty}{L}\, S_n\cdot T_n = \underset{n\to\infty}{L}\, S_n\cdot \underset{n\to\infty}{L}\, T_n = AB.$$

It can be shown that if this product series converges at all, it converges to the limit $A \cdot B$.[1] The difficulty is that it does not always converge as the following example shows.

**Ex. 6.** Consider the product of the series $\Sigma(-1)^{n-1}\frac{1}{\sqrt{n}}$ by itself.

The given series

$$\Sigma(-1)^{n-1}\frac{1}{\sqrt{n}} = 1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \cdot \cdot \cdot ,$$

converges, since the terms are alternately positive and negative and the condition for the convergence of such a series exists; namely,

$$\underset{n\to\infty}{L}\, u_n = \underset{n\to\infty}{L}\, \frac{1}{\sqrt{n}} = 0.$$

However, the product of this series by itself gives the series

$$1 - \frac{2}{\sqrt{2}} + \left(\frac{2}{\sqrt{3}} + \frac{1}{\sqrt{4}}\right) - \left(\frac{2}{\sqrt{4}} + \frac{2}{\sqrt{6}}\right) + \cdot \cdot \cdot$$
$$\pm\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{2(n-1)}} + \frac{1}{\sqrt{3(n-2)}} + \cdot \cdot \cdot + \frac{1}{\sqrt{2(n-1)}} + \frac{1}{\sqrt{n}}\right) \mp \cdot \cdot \cdot ,$$

which is a divergent series. In fact, since

$$k(n-k+1) = \frac{1}{4}(n+1)^2 - \left\{\frac{1}{2}(n+1) - k\right\}^2 \leqq \frac{1}{4}(n+1)^2,$$

we have

$$\left|\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{2(n-1)}} + \frac{1}{\sqrt{3(n-2)}} + \cdot \cdot \cdot + \frac{1}{\sqrt{2(n-1)}} + \frac{1}{\sqrt{n}}\right| > \frac{1}{\frac{1}{2}(n+1)} \cdot n = \frac{2}{1+\frac{1}{n}}.$$

As all the terms of the product series are equal to or greater than 1, it follows that the series is divergent.

The following theorem due to Mertens[2] gives a useful criterion for the multiplication of series.

THEOREM IV. *Let the series $\Sigma\, u_n$, $\Sigma\, v_n$ converge to the limits $A$, $B$, respectively. If one of these series converges absolutely, then*

$$A \cdot B = \Sigma\, u_n \cdot \Sigma\, v_n = \Sigma\, p_n,$$

*where* $$p_n = u_1v_n + u_2v_{n-1} + u_3v_{n-2} + \cdot \cdot \cdot + u_nv_1.$$

---

[1] See Abel, *Crelle's Journal*, Vol. 1 (1827). The older mathematicians assumed that the product of two infinite series follows the rule for the multiplication of two polynomials without any conditions as to convergence. Thus De Moivre (1697) extended the binomial theorem to infinite series without raising the question of convergence. The first to call attention to the necessity of considering the convergence of the series was Cauchy, who showed that two absolutely convergent series may be multiplied according to the method suggested and that the resulting series is absolutely convergent.

[2] See *Crelle's Journal* (1875) Vol. 79, p. 182.

Let us assume that $\Sigma\, v_n$ converges absolutely, and denote by $P_n$ the sum of those terms $u_r \cdot v_s$ of the product such that

$$r + s \leqq n + 1;$$

that is, $$P_n = p_1 + p_2 + \cdot \cdot \cdot + p_n .$$

Hence, we have

$$P_n = (u_1 + u_2 + \cdot \cdot \cdot + u_n)(v_1 + v_2 + \cdot \cdot \cdot + v_n) - v_2 u_n - v_3(u_n + u_{n-1}) - \cdot \cdot \cdot - v_n(u_2 + u_3 + \cdot \cdot \cdot + u_n). \quad (5)$$

Let $$S_n = u_1 + u_2 + \cdot \cdot \cdot + u_n,$$

$$T_n = v_1 + v_2 + \cdot \cdot \cdot + v_n.$$

Subtracting $AB$ from both members of (5), we have upon taking the absolute value

$$|P_n - AB| \leqq |S_n \cdot T_n - AB| + |v_2|\ |u_n| + |v_3|\ |u_{n-1} + u_n| + \cdot \cdot \cdot + |v_n|\ |u_2 + u_3 + \cdot \cdot \cdot + u_n|. \quad (6)$$

Let $k$ be some integer between 1 and $n$, say $\frac{n}{2}$ if $n$ is even and $\frac{n-1}{2}$ if $n$ is odd. Denote by $\delta$ the largest of the positive numbers

$$|u_n|,\ |u_n + u_{n-1}|,\ \cdot \cdot \cdot,\ |u_n + \cdot \cdot \cdot + u_{n-k}|,$$

and by $\gamma$ the largest of the numbers

$$|u_n + \cdot \cdot \cdot + u_{n-k-1}|,\ \cdot \cdot \cdot,\ |u_n + u_{n-1} + \cdot \cdot \cdot + u_2|.$$

We then have from (6)

$$|P_n - AB| \leqq |S_n T_n - AB| + \delta\{|v_2| + |v_3| + \cdot \cdot \cdot + |v_{k+2}|\} + \gamma\{|v_{k+3}| + \cdot \cdot \cdot + |v_n|\}. \quad (7)$$

As $n$ increases indefinitely, $\delta$ and $\{|v_{k+3}| + \cdot \cdot \cdot + |v_n|\}$ decrease and each approaches the value zero, while $\gamma$ and $\{|v_2| + \cdot \cdot \cdot + |v_{k+2}|\}$ remain finite. Consequently, as $n$ becomes infinite each term in the right-hand member of (7) has the limiting value zero. Hence, we have

$$\underset{n \to \infty}{L}\ P_n = AB,$$

as was to be shown.

Since the series $\Sigma\, u_n$ and $\Sigma\, v_n$ may be multiplied when either of them is absolutely convergent, it follows, of course, that the same is true when both converge absolutely. In the latter case, one may rearrange the terms in any order at pleasure, and we may take the product by grouping together, as we have done, the terms in the successive diagonals of (3), or by taking them in any other order at pleasure which enables us to include in the result all the terms of the product.

**Ex. 7.** Find the product of the two series

$$1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdot \cdot \cdot ,$$

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdot \cdot \cdot .$$

Both series converge, the first to the value $e$ and the second to the value log 2. The first converges absolutely and the second conditionally. However, by the foregoing theorem the product of the two series must converge to the limiting value of $e \log 2$.

It may be shown that the product of two conditionally convergent series will not always be an absolutely convergent series, although, as Abel's investigation already quoted shows, two such series may be multiplied by the foregoing method provided the resulting series converges. However, a conditionally convergent series or even a divergent series multiplied by a conditionally convergent series will under certain conditions give a product that is absolutely convergent.[1]

**59. Uniform convergence of a series.** Suppose we have given a series of functions

$$\Sigma\, u_n(x) = u_1(x) + u_2(x) + \cdot \cdot \cdot + u_n(x) + \cdot \cdot \cdot ,$$

which converges for each value of $x$ in a given interval $(a, b)$. In this interval the series defines a function of $x$ and we may write

$$f(x) = \Sigma\, u_n(x).$$

Not all of the functional properties of the individual terms $u_n(x)$ are necessarily valid for the function $f(x)$ defined by the series. For example, each $u_n(x)$ may be continuous without $f(x)$ being continuous. As an illustration, we have the following example.

---

[1] See Pringsheim, *Math. Ann.*, Vol. 21 (1882), pp. 327–378; also *Trans. Amer. Math. Soc.*, Vol. 2 (1901), pp. 404–412. He also sets up a sufficient condition under which two conditionally convergent series may be multiplied if the terms of one of the series can be arranged in groups each containing a finite number of terms so that this series of groups converges absolutely. Voss has developed a necessary and sufficient condition for the multiplication of two such series where there are but two terms in each group. [See *Math. Ann.*, Vol. 24 (1884), pp. 42–47.] Cajori has extended the results of Voss to the case where the number of terms in each group is variable [see *Bulletin of the Amer. Math. Soc.*, 2$^{d}$ series, Vol. 1 (1895), p. 180; Vol. IX (1903), p. 188; *Trans. Amer. Math. Soc.*, Vol. 2 (1901), pp. 25–36].

**Ex. 1.** Given the series for which $S_n(x) = x^{\frac{1}{2n-1}}$, where $-1 \leqq x \leqq 1$. Test the limiting function $f(x)$ for points of discontinuity.

For all finite values of $n$, $S_n(x)$ is a continuous function of $x$. However,

$$f(x) = \underset{n\to\infty}{L}\ S_n(x)$$

is not continuous at $x = 0$; for, we have

$$\begin{aligned} f(x) &= -1\ , \text{ for } x < 0, \\ &= \phantom{-}0\ , \text{ for } x = 0, \\ &= \phantom{-}1\ , \text{ for } x > 0. \end{aligned}$$

Likewise, it can not be assumed that because each term of a series can be integrated the sum of the integrals thus formed is the integral of the function defined by the series, even though the series of integrals converges. A similar statement can be made with reference to the differentiation of a series term by term. Evidently, then, one must know more about a series than the mere fact that it converges in order to answer some of the more fundamental questions about the operations upon it by the ordinary processes of the calculus. A sufficient condition for many of these operations is given by what is known as the uniform convergence of the series.

When we say that a series $\Sigma\, u_n(x)$ converges for all values of $x$ in an interval $(a, b)$ we mean that for each value of $x$, say $x_0$, in that interval the series of constant values $\Sigma\, u_n(x_0)$ is convergent; that is, we have

$$\underset{n\to\infty}{L}\ S_n(x_0) = A \equiv f(x_0).$$

Written as an inequality, this relation means that for a previously assigned, arbitrarily small, positive value $\epsilon$, we have for all values of $n$ greater than some integer $m$

$$|\, S_n(x_0) - f\,(x_0)\,| < \epsilon,\ n > m. \tag{1}$$

The value of $m$ depends both upon the value assigned to $\epsilon$ and the choice of the particular value of $x$. For any given $\epsilon$, the value of $m$ is a function of $x$, and to each value of $x$ in the given interval there corresponds a definite value of $m$. The aggregate of all these values of $m$ as $x$ takes all values in $(a, b)$, may have a finite upper bound. If this is the case, we say that the given series **converges uniformly** in the interval $(a, b)$; otherwise, the convergence is said to be non-uniform.[1] In the case of uniform convergence, it is then possible for an arbitrarily chosen value of $\epsilon$ to select a value of $m > 0$ which does not depend upon $x$ and yet satisfies the conditions imposed upon it by (1); that is, for a given $\epsilon$ it is possible to so choose the value of $m$ that it answers uniformly for all values of $x$ in the interval $(a, b)$.

---

[1] The notion of the uniform convergence of an infinite series was first developed by Seidel (1848) in a paper published in *Abh. Akad. München*, Vol. $5_2$, pp. 379–393, and by Stokes (1849) in a paper published in the *Trans. Cambridge Philos. Society*, Vol. $8_5$, pp. 533–583.

As we have seen (Fig. 40), ordinary convergence of a series means geometrically that for a given $\epsilon$ and a given value of $x$, the values of

$$y = S_n(x_0),\ n > m(x_0)$$

lie within a particular rectangle situated in a plane through the line $x = x_0$ and parallel to the coördinate plane $\left(\frac{1}{n}, y\right)$. Uniform convergence means that for a given value of $\epsilon$ all of these rectangles may be taken to be the same size whatever value of $x$ may be selected. The curves representing

$$y = S_n(x)$$

are known as the **approximation curves.** Denoting the upper bound of $m(x)$ by $m_1$, it follows in the case of uniform convergence that the approximation curves for $n > m_1$ all project into curves in the $(XY)$-plane which lie within a strip bounded by the curves

$$y = f(x) \pm \epsilon,$$

as indicated in Fig. 43. Instead of considering the curves $y = S_n(x)$ in three-dimensional space, it is often more convenient to consider only the projections of these curves upon the $(X\,Y)$-plane. If this is done, the representation reduces to two dimensions as illustrated by the following example.

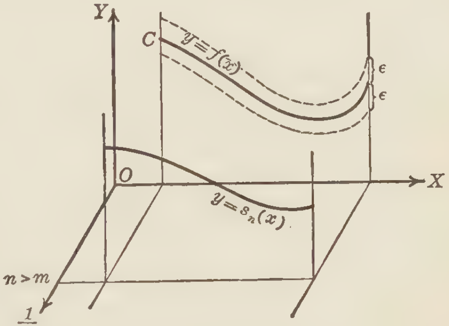


FIG. 43.

**Ex. 2.** Draw the approximation curves and show the uniform convergence of the series

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \cdot \cdot \cdot + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + \cdot \cdot \cdot .$$

This series converges in the interval (0, 1) and defines the function

$$y = \text{arc tan } x.$$

Several of the approximation curves

$$y = S_1(x) = x,$$
$$y = S_2(x) = x - \frac{x^3}{3},$$
$$y = S_3(x) = x - \frac{x^3}{3} + \frac{x^5}{5},$$
$$\cdots\cdots\cdots$$

are shown in Fig. 44. It will be observed that any two consecutive approximation curves lie on opposite sides of the curve $y = f(x)$, and moreover each lies closer to this curve than any of the preceding. For $n$ sufficiently large, all of the approximation curves lie wholly within the $\epsilon$ strip indicated. Consequently, the given series converges uniformly in the interval (0, 1).

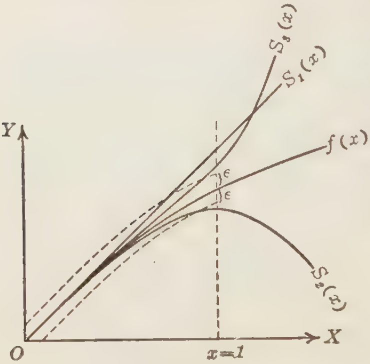


FIG. 44.

**Ex. 3.** Draw the approximation curves and show the non-uniform convergence of the series for which

$$S_n(x) = \frac{n^2x}{1 + n^3x^2}, \quad 0 \leqq x \leqq 1.$$

The series converges for all values of $x$ in the given interval; for, we have

$$f(x) = \underset{n\to\infty}{L}\ S_n(x) = \underset{n\to\infty}{L}\ \frac{n^2x}{1 + n^3x^2} = 0.$$

The function $f(x)$ is then represented by a segment of the $X$-axis. Two of the approximation curves are shown in Fig. 45.

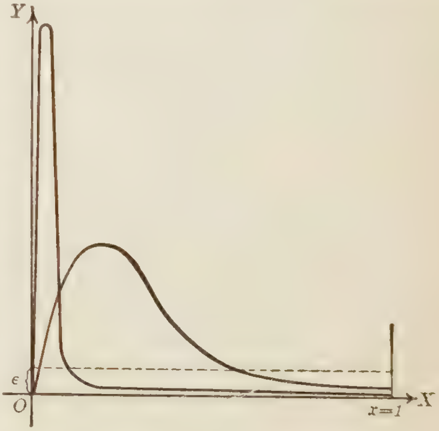


FIG. 45.

It will be observed that the approximation curves in this case have a peak which grows higher and sharper as $n$ increases. It is evident that these curves can never lie wholly within the $\epsilon$ strip, however large $n$ may be chosen; for, the peak always extends beyond this strip. The peak, however, is situated closer and closer to the origin as $n$ increases, and, should we exclude an arbitrarily small interval including the origin, the value of $n$ can be taken so large that the peak would, from that point on, always lie within this small interval. Consequently, while the given series does not converge uniformly in the interval (0, 1), it does converge uniformly in any interval $(\eta, 1)$ where $\eta$ is positive and as small as we please to make it.

Thus far we have considered the uniform convergence of an infinite series with respect to an interval. One may generalize this notion and discuss **uniform convergence with respect to a set of points.**

One may likewise consider the **uniform convergence of a series at a single point.** If we have given the relation

$$\underset{\substack{n\to\infty \\ x\to x_0}}{L} R_n(x) = 0,$$

or what is the same thing, the relation

$$\underset{\substack{n\to\infty \\ x\to x_0}}{L} S_n(x) = f(x_0),$$

then the point $x = x_0$ is sometimes spoken of as a point of uniform convergence of the given series.

The following theorem gives us a convenient test for uniform convergence, known as the **Weierstrass test.**

THEOREM I. *A given series $\Sigma\, u_n(x)$ will converge uniformly in the interval $(a, b)$ if the series $\Sigma\, M_n$ of positive terms, where $|\, u_n(x)\,| \leqq M_n$ for all values of $x$ in $(a, b)$, is convergent.*

Put $$R_n(x) = u_n(x) + u_{n+1}(x) + \cdot\ \cdot\ \cdot\ .$$

We have from the given hypothesis

$$|\, R_n(x)\,| \leqq M_n + M_{n+1} + \cdot\ \cdot\ \cdot\ .$$

But as the series $\Sigma\, M_n$ converges, the right-hand member of this inequality can be made less than any arbitrarily small value $\epsilon$ by taking $n$ sufficiently large, say $n > m$. We have then

$$|\, R_n(x)\,| \equiv |\, S_n(x) - f(x)\,| < \epsilon,$$

for all values of $x$ in $(a, b)$ and for $n > m$, where $m$ is independent of $x$. This is nothing else than definition of uniform convergence.

**Ex. 4.** Test for uniform convergence the series

$$\frac{4}{\pi}\left\{\sin x - \frac{\sin 3x}{3^2} + \frac{\sin 5x}{5^2} - \cdot\ \cdot\ \cdot\right\},\ -\pi \leqq x \leqq \pi.$$

Forming the series of greatest numerical values of the terms of the given series, we have for all values of $x$ in question

$$\Sigma\, M_n = \frac{4}{\pi}\left\{1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdot\ \cdot\ \cdot\right\}.$$

This latter series is convergent, since its terms are less than the corresponding terms of the known convergent series $\Sigma \frac{1}{n^2}\cdot$ Hence, $\Sigma M_n$ converges and the given series converges uniformly. Some of the approximation curves are shown in Fig. 46.

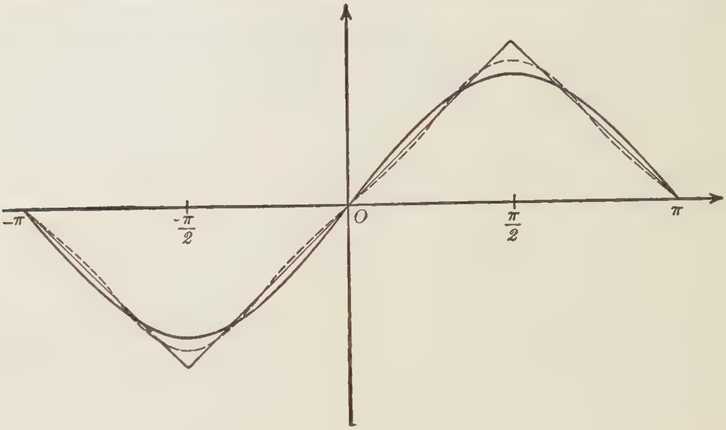


FIG. 46

It will be observed that $S_n(x)$ is in reality a function of the two independent variables $x$ and $n$. Consequently, those properties already developed in Art. 26 with reference to the uniform convergence of a function $f(x, y)$ to the limiting function $f(x, y_0)$ apply directly to the uniform convergence of an infinite series. For example, if the terms $u_n(x)$ are continuous functions of $x$, then $S_n(x)$ is continuous and we have from Theorem VII of Art. 26 the following theorem.

THEOREM II. *Given the series* $\Sigma\, u_n(x)$, *where* $u_n(x)$ *are continuous functions of* $x$. *A necessary and sufficient condition that* $\Sigma\, u_n(x)$ *converges uniformly in the interval* $(a, b)$ *is that the simultaneous limit* $\underset{\substack{x\to x_0\\ n\to\infty}}{L} S_n(x)$ *exists for each point* $x_0$ *in the closed interval* $(a, b)$.

**Ex. 5.** Given the series the sum of whose first $n$ terms is $S_n(x) = \frac{n^2x}{1+n^4x^2}\cdot$ Show that it converges non-uniformly in the interval $0 \leqq x \leqq 1$.

For all finite values of $n$, $S_n(x)$ is a continuous function of $x$ in the given interval. The series converges for all values of $x$ in the given interval, for if $x_0$ is any value of $x$ in this interval, we have

$$\underset{n\to\infty}{L} S_n(x_0) = \underset{n\to\infty}{L} \frac{n^2x_0}{1+n^4x_0^2} = 0 = f(x_0).$$

However, the simultaneous limit

$$\underset{\substack{x\to 0\\ n\to\infty}}{L} S_n(x) = \underset{\substack{x\to 0\\ n\to\infty}}{L} \frac{n^2x}{1+n^4x^2}$$

does not exist. For, if we let $n$ become infinite and then allow $x$ to approach zero we have the limiting value zero, while along the path $x = \frac{1}{n^2}$ we obtain the limiting value $\frac{1}{2}$. Since the simultaneous limit does not exist at one point of the closed interval $(0, 1)$, then by the foregoing theorem the convergence is non-uniform in the given interval. If any interval inclosing the origin is omitted; for example, if we consider the interval $(\epsilon, 1)$, where $\epsilon > 0$, the series is uniformly convergent in that interval.

It follows from the foregoing theorem that if a series converges uniformly, one may interchange the order of the limits $\underset{n \to \infty}{L}$ $\underset{x \to x_0}{L}$ for any $x_0$ in the interval of uniform convergence. Thus we may write

$$\underset{x \to x_0}{L} \ \underset{n \to \infty}{L} \ S_n(x) = \underset{n \to \infty}{L} \ \underset{x \to x_0}{L} \ S_n(x).$$

Theorem VIII of Art. 26, stated for the special case of an infinite series, gives the following theorem.

THEOREM III. *If $S_n(x)$ has a derivative $S_n'(x)$ for all finite values of $n$ and for all values of $x$ in the closed interval $(a, b)$, and if for all such values of $x$ and $n$, we have*

$$| S_n'(x) | < M,$$

*where $M$ is a finite number independent of $x$, then the given series converges uniformly in $(a, b)$.*

**Ex. 6.** Test for uniform convergence the series

$$1 + x + x^2 + x^3 + \cdot \cdot \cdot + x^n + \cdot \cdot \cdot, \ |x| < \frac{1}{2}.$$

We have
$$S_n(x) = \frac{1 - x^n}{1 - x},$$

whence for all values of $x$ in the given interval and for all values of $n$ sufficiently large we have

$$\left| S_n'(x) \right| = \left| \frac{1 - nx^{n-1} + (n-1)x^n}{(1 - x)^2} \right| < 4.$$

Consequently, in the interval $(-\frac{1}{2}, \frac{1}{2})$ the given series converges uniformly.

From the discussion of the properties of the limiting function in the case of two variables (Theorem XI, Art. 28), it will be seen that for the special case of infinite series we may state the following theorem.

THEOREM IV. *Let the series $\Sigma\, u_n(x)$ converge in the interval $(a, b)$ and suppose $u_n(x)$ to be a continuous function of $x$ in this interval for all values of $n$. Then the points of uniform convergence of $\Sigma\, u_n(x)$ form a set which is everywhere dense in $(a, b)$ and has the cardinal number of the continuum.*

Some of the consequences of uniform convergence follow as special applications of the results of the discussion in Art. 26 of the uniform convergence of a function $f(x, y)$ to the limiting function $f(x, y_0)$. For example, from Theorem IX, we have the following theorem.

THEOREM V. *If the terms of an infinite series are continuous functions of $x$ in a given interval $(a, b)$ and the series converges uniformly in this interval, then the function defined by the series is also continuous in the same interval.*

It may also be shown that every continuous function may be represented by a series of polynomials. Indeed, if the interval is closed, it may be shown that this series of polynomials converges uniformly and absolutely.[1]

The condition set forth in the foregoing theorem is a sufficient rather than a necessary one. The following example furnishes an illustration.

**Ex. 7.** Given the series $\Sigma\, u_n$, for which

$$S_n(x) = \frac{nx}{1 + n^2x^2}, \quad 0 \leqq x \leqq 1.$$

Test the uniform convergence of the series and the continuity of the function defined by it.

In the given interval $S_n(x)$ is continuous, but the limit $\underset{\substack{n\to\infty \\ x\to 0}}{L} S_n(x)$ does not exist. Hence, by Theorem II the series does not converge uniformly, a fact which may also be seen from the approximation curves, which are the same as those given in Fig. 22. However, the series defines the function

$$f(x) = \underset{n\to\infty}{L} \frac{nx}{1 + n^2x^2} = 0,$$

and hence is continuous in the given interval.

If we assume merely that the given series of continuous functions converges, it then follows from Theorem IV that the points of uniform convergence are at least everywhere dense in the given interval and form a set having the cardinal number of the continuum. Since the limit $\underset{n\to\infty}{L} S_n(x)$ exists for each $x$ in the interval of convergence, it follows from Theorem V, Art. 24 that at the points of uniform convergence, that is where $\underset{\substack{n\to\infty\\ x\to x_0}}{L} S_n(x)$ exists, the limit $\underset{x\to x_0}{L}\ \underset{n\to\infty}{L} S_n(x)$ also exists and is equal to $f(x_0)$. In other words, we have

---

[1] Due to Weierstrass, *Berliner Sitzungsberichte* (1885); see also Borel's *Leçons sur les fonctions de variables reelles*, Chapters IV, V, for discussion of various proofs and citation of literature.

$$\underset{x\to x_0}{L} f(x) = f(x_0).$$

Consequently, at these points the limiting function $f(x)$ is continuous and we have the following theorem.

THEOREM VI. *If the series $\Sigma\, u_n(x)$ converges in the interval $(a, b)$ and if each $u_n(x)$ is continuous, then the limiting function $f(x)$ is at most pointwise discontinuous. Moreover, the points of continuity form a set having the cardinal number of the continuum.*

It must not be assumed that because the terms of a series are discontinuous it can not converge uniformly. The following is such a series and defines a function known as the Riemann function.[1]

**Ex. 8.** Denote by $(x)$ the positive or negative excess of $x$ over the nearest larger or smaller integer. If $x$ is midway between two integers, let $(x)$ be zero. In any finite interval, test the uniform convergence of the series

$$\Sigma\frac{(nx)}{n^2} = \frac{(x)}{1} + \frac{(2x)}{2^2} + \frac{(3x)}{3^2} + \cdot\ \cdot\ \cdot + \frac{(nx)}{n^2} + \cdot\ \cdot\ \cdot\ .$$

It will be seen that $\frac{1}{2}$ is the least upper bound of $(nx)$. Hence, the Weierstrass test for uniform convergence applies; for, the series of greatest numerical values of the individual terms is less term by term than the convergent series $\Sigma\frac{1}{n^2}$. Each term of the series is discontinuous. For example, the first term is discontinuous at the points

$$\tfrac{1}{2},\ 1\tfrac{1}{2},\ 2\tfrac{1}{2},\ \cdot\ \cdot\ \cdot\ .$$

The second term is discontinuous at the points

$$\tfrac{1}{4},\ \tfrac{2}{4},\ \tfrac{3}{4},\ \cdot\ \cdot\ \cdot\ .$$

All the points of discontinuity of the various terms are points of discontinuity of $f(x)$. In fact, every value of $x = \frac{p}{2m}$, where $p$ is prime to $2m$, is a point of discontinuity of $f(x)$. The totality of these points forms a set of content zero.

### 60. Simply-uniform convergence. Quasi-uniform convergence.

From the discussions in the previous article, it is evident that while ordinary uniform convergence gives a sufficient condition for the application of some of the fundamental operations of the calculus to infinite series, a more restrictive condition is desirable. For this purpose Dini[2] introduced what he called **simply-uniform conver-**

---

[1] See *Gesammelte Werke*, p. 228.

[2] See *Fundamenti per la teorica delle funzioni di variabili reali* (1878), p. 103, or German Ed., *Gründlagen für eine Theorie der Functionen*, p. 137.

gence. A series $\Sigma\, u_n(x)$ defining the function $f(x)$ is said to converge simply-uniformly if for each integer $m$ however large there exists at least one value of $n$ greater than $m$ for which we have

$$|\, R_n(x)\,| \equiv |\, S_n(x) - f(x)\,| < \epsilon, \tag{1}$$

for all values of $x$ in an interval $(a, b)$.

For example, suppose we have given a series whose terms are formed in accordance with the law[1]

$$u_{2n-1} = \frac{x}{nx^2 + (1 - nx)^2}, \quad u_{2n} = \frac{-x}{(n+1)x^2 + [1 - (n+1)x]^2},$$

where $n = 1, 2, 3, \cdots$. We have then the series

$$\frac{x}{x^2 + (1 - x)^2} - \frac{x}{2x^2 + (1 - 2x)^2} + \frac{x}{2x^2 + (1 - 2x)^2} - \cdots.$$

This series converges for all values of $x$ in the interval $-1 < x \leqq 1$ and defines the function

$$f(x) = \frac{x}{x^2 + (1 - x)^2}.$$

For an odd number of terms, we have

$$|\, S_{2n-1}(x) - f(x)\,| = 0,$$

and hence the inequality (1) is satisfied for those values of $n$, and the series converges simply-uniformly. For an even number of terms, we have

$$|\, R_{2n}(x)\,| = |\, S_{2n}(x) - f(x)\,| = \frac{x}{nx^2 + (1 - nx)^2}.$$

In order that the given series shall converge uniformly in the ordinary sense, we must have

$$|\, R_n(x)\,| < \epsilon,$$

for all values of $n$ greater than some value $m$ independent of $x$ and for all values of $x$ in the interval in question. That this is not true, if $n$ is an even integer, may be shown by taking the limit along the line

$$x = \frac{1}{n},$$

by which we have

$$\underset{n\to\infty}{L}\, |\, R_n(x)\,| = \frac{\frac{1}{n}}{n\cdot\frac{1}{n^2} + (1 - 1)^2} = \underset{n\to\infty}{L}\, \frac{\frac{1}{n}}{\frac{1}{n}} = 1.$$

---

[1] See Tannery, *Théorie des fonctions*, p. 134.

While the series can not therefore converge uniformly in the ordinary sense, it does converge simply-uniformly as we have seen. It follows at once from the foregoing definition that every series which converges uniformly must also converge simply-uniformly, although the converse is not necessarily true.

It is to be noted, however, that by the proper grouping of the terms of a simply-uniform convergent series one may always form a new series which converges uniformly. For, let the given simply-uniform convergent series be

$$u_1(x) + u_2(x) + u_3(x) + \cdot \cdot \cdot + u_n(x) + \cdot \cdot \cdot .$$

Since it converges simply-uniformly, there must exist a sequence of integral values $\{n_k\}$ dense at infinity, such that for each value of $n_k$ greater than some integer $m$, we have for all values of $x$ in the given interval $(a, b)$

$$| S_{n_k}(x) - f(x) | < \epsilon,$$

where $\epsilon$ is arbitrarily small. However, $S_{n_k}(x)$ may be written in the form

$$S_{n_k}(x) = S_{n_1}(x) + \{S_{n_2}(x) - S_{n_1}(x)\} + \cdot \cdot \cdot \\ + \{S_{n_k}(x) - S_{n_{k-1}}(x)\}. \quad (2)$$

We may regard each of the differences $\{S_{n_k}(x) - S_{n_{k-1}}(x)\}$ as a term of a new series and denote the sum of the first $k$ terms given in (2) by $S'_{n_k}(x)$. We have then for all values of $x$ in $(a, b)$ and for all values of $n_k > m$

$$| S'_{n_k}(x) - f(x) | < \epsilon,$$

which is none other than the condition that the series

$$S_{n_1}(x) + \{S_{n_2}(x) - S_{n_1}(x)\} + \cdot \cdot \cdot + \{S_{n_k}(x) - S_{n_{k-1}}(x)\} + \cdot \cdot \cdot$$

converges uniformly in the interval $(a, b)$. If in the grouping of terms the order of the original terms has not been changed, the new series converges to the same limiting function as the original series. If the given series converges absolutely for each value of $x$, the same is true even though the order is disturbed by the grouping of the terms.

Another special form of convergence has been introduced by Arzelà, called **quasi-uniform convergence.**[1] This form of convergence differs from ordinary uniform convergence and from simply-uniform convergence in that the inequality (1) does not necessarily hold for any single value of $n$ and all values of $x$ in the given interval. Quasi-uniform convergence of an infinite series may be defined as follows.

---

[1] See *Bologna Rendiconto* (Apr. 1884); also *ibid.* (Nov. 1902).

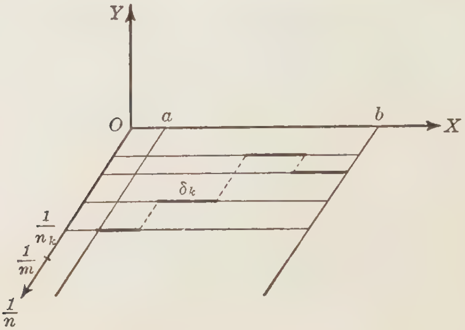


FIG. 47.

If for each value of $n$ from a certain point on, say $n > m$, there exists a finite number of values $n_k > n$, such that for an arbitrarily small value of $\epsilon$ there exists upon each line (Fig. 47) through $\frac{1}{n_k}$ and parallel to the $X$-axis an interval of length $\delta_k$, which may vary with $x$, such that the sum of the projections of these intervals upon the $X$-axis just fills the given interval $(a, b)$, and if for each $x$ on the broken line thus formed from these intervals we have

$$| R_{n_k}(x) | \equiv | S_{n_k}(x) - f(x) | < \epsilon, \tag{3}$$

then the given series is said to converge quasi-uniformly in $(a, b)$.

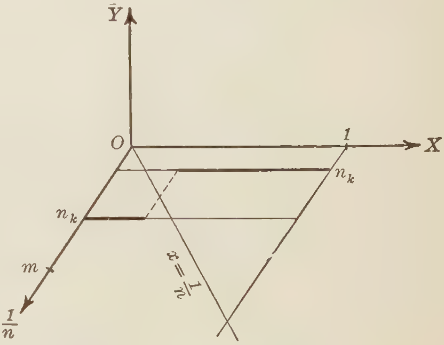


FIG. 48.

It follows from this definition that any series which converges uniformly or simply-uniformly converges also quasi-uniformly. The converse, however, is not true. Thus in Ex. 7, Art. 59, it was seen that the series for which we have

$$S_n = \frac{nx}{1 + n^2x^2}$$

does not converge uniformly in the interval (0, 1). The reason for the non-uniformity of the convergence is the presence of a peak of the approximation curve in the neighborhood of the origin. Along the line $x = \frac{1}{n}$, the value of

$$y = S_n(x),$$

is always $\frac{1}{2}$. However, for a finite number of values of $n$, say $n = n_k$, $n_k > m$, there always exist intervals on the lines

$$\frac{1}{n} = \frac{1}{n_k},$$

where $n_k$ varies with $x$ as indicated in Fig. 48, so that for every $x$ in on the broken line made up of these intervals the inequality (3) is valid, and hence the series converges quasi-uniformly.

If a series converges quasi-uniformly in a given interval, except for a finite number of subintervals the sum of whose lengths may be made arbitrarily small, then the series is said to converge **quasi-uniformly in general** in the given interval.

The question of quasi-uniform convergence is of special importance in considering such questions as the integration of a series term by term, the continuity of the function defined by a series, etc. As will be seen at once, the quasi-uniform convergence of a series is a special case of the quasi-uniform convergence of a function $f(x, y)$ to the limiting function $f(x, y_0)$ already discussed in Art. 26. Consequently, the theorems there developed can be restated without proof for the case of infinite series. For example, we have from Theorem XI of that article the following theorem.

THEOREM I. *Let the series* $\Sigma\, u_n(x)$ *of continuous functions* $u_n(x)$ *converge in the closed interval* $(a, b)$. *A necessary and sufficient condition that the function* $f(x)$ *defined by this series is continuous in* $(a, b)$ *is that the given series converges quasi-uniformly in this interval.*

That a function defined by a series of continuous functions is not necessarily continuous has already been pointed out in connection with Ex. 1 of Art. 59. The function there defined by a series of continuous functions is discontinuous at the point $x = 0$. It follows from the discussions in Art. 28 of the limiting function $f(x, y_0)$ that if a function $f(x)$ is defined by a series of continuous functions, then $f(x)$ can at most be point-wise discontinuous. Moreover, it may be shown that the converse of this statement is also true; namely, every point-wise discontinuous function can be defined by a convergent series of continuous functions.[1] We are therefore led to the conclusion that a necessary and sufficient condition that a function defined by a series of continuous functions be continuous is that the function is at most point-wise discontinuous.

---

[1] See Baire's *Leçons sur les fonctions discontinuous*, p. 124.

If one is interested in the continuity of the function $f(x)$ at a single point $x_0$ rather than in an interval, then one may set up a necessary and sufficient condition by making use of Theorem X Art. 26. Stated for the case of an infinite series, it is as follows.

THEOREM II. *Given a series* $\Sigma\, u_n(x)$ *of continuous functions* $u_n(x)$ *which converges in the closed interval* $(a, b)$ *and defines the function* $f(x)$. *In order that* $f(x)$ *be continuous at any point* $x_0$ *of this interval, it is necessary and sufficient that for every value of* $n$ *however large there exists an integer* $n_k > n$ *and likewise an interval containing the point* $x_0$ *as an inner point, which interval may vary in extent with* $n_k$, *such that for all values of* $x$ *in this interval we have*

$$| f(x) - S_{n_k}(x) | < \epsilon,$$

*where* $\epsilon$ *is an arbitrarily small positive number.*

**61. Integration of series.** We shall now inquire into the conditions under which one may integrate a series term by term and thus obtain the integral of the function defined by the series. By the integral of the function thus defined, we understand the integral of the limit of $S_n(x)$ as $n$ becomes infinite; that is, we have

$$\int_a^b f(x)dx = \int_a^b \underset{n\to\infty}{L}\, S_n(x)\, dx.$$

On the other hand, the determination of $\int_a^b f(x)dx$ by the integration of a series term by term implies that we have

$$\int_a^b f(x)dx = \underset{n\to\infty}{L} \int_a^b S_n(x)dx.$$

The question is then as to the conditions under which one may write

$$\int_a^b \underset{n\to\infty}{L}\, S_n(x)dx = \underset{n\to\infty}{L} \int_a^b S_n(x)dx.$$

That this is not always possible, is shown by the following example.[1]

**Ex. 9.** Given the series, the sum of whose first $n$ terms is $nxe^{-nx^2}$. Show that this series can not be integrated term by term and thus obtain the integral of the function defined by the series.

We have $\int_0^1 \underset{n\to\infty}{L}\, S_n(x)dx = \int_0^1 \underset{n\to\infty}{L}\, nxe^{-nx^2}\, dx = \int_0^1 0\cdot dx = 0.$

---

[1] For other illustrations, see Hardy, *Messenger of Mathematics*, Vol. 44(1915), p. 145.

On the other hand, we have

$$\underset{n\to\infty}{L}\int_0^1 S_n(x)dx = \underset{n\to\infty}{L}\int_0^1 nxe^{-nx^2}\,dx = \underset{n\to\infty}{L}\,\frac{1-e^{-n}}{2} = \frac{1}{2}.$$

As these two results are different, the given series if integrated term by term does not give the integral of the function defined by the series.

The problem to be considered may be regarded as a special case of those questions already discussed in connection with functions of two independent variables. For, we need only to replace the variable $y$ in $f(x, y)$ by $\frac{1}{n}$ and consider the resulting function for integral values of $n$. Similarly, from those theorems concerning sequences of functions $\{f_n(x)\}$ we have corresponding theorems for infinite series; for, we can always replace $f_n(x)$ by $S_n(x)$ and restate the theorems.

If each of the terms of a given series $\Sigma\, u_n(x)$ is integrable, one may question whether the function $f(x)$ defined by the series is also integrable. The answer to this question is given by Theorem II, Art. 46, and may be stated for infinite series as follows:

THEOREM I. *Given a bounded function $f(x)$ defined in the interval $(a, b)$ by the infinite series $\Sigma\, u_n(x)$, where $u_n(x)$ is bounded and integrable in the given interval. A necessary and sufficient condition that $f(x)$ is integrable in $(a, b)$ is that the series $\Sigma\, u_n(x)$ converges quasi-uniformly in general in $(a, b)$.*

We have seen that the indefinite integral of an integrable function is a continuous function throughout the interval of integration. Under certain conditions this is also true of the sum of the indefinite integrals of the terms of a series. In other words, it may happen that $\phi(x)$ defined by the relation

$$\phi(x) = \underset{n\to\infty}{L}\int_a^x S_n(x)dx,$$

where $S_n(x)$ is integrable, may be a continuous function.

One might infer that under these circumstances the given series could be integrated term by term and the series of integrals thus obtained represent the integral of the function defined by the series. That this is not always the case is shown by the following illustration.[1]

---

[1] See Vitali, *Rendiconti del Circolo Mathematico di Palermo*, Vol. 23 (1907), p. 155; for another illustration of same kind, see Osgood, *Amer. Jour. of Math.*, Vol. 19 (1897), p. 167.

Let $S_n(x)$ be defined in the given interval (0, 1) as follows:

$S_1(x) = 1$ in the interval (0, 1),

$S_2(x) = 4$ in the intervals $(0, \frac{1}{8})$, $(\frac{1}{2}, \frac{5}{8})$ and otherwise zero,

$S_3(x) = 9$ in the intervals $(0, \frac{1}{27})$, $(\frac{1}{3}, \frac{10}{27})$, $(\frac{2}{3}, \frac{19}{27})$ and otherwise zero,

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

$S_n(x) = n^2$ in the intervals $\left(\frac{k}{n}, \frac{k}{n} + \frac{1}{n^3}\right)$, $k = 0, 1, 2, \cdots (n-1)$ and otherwise zero.

It will be observed that the number of intervals where $S_n(x)$ is different from zero is always the same as the subscript of $S_n(x)$ and the value of $S_n(x)$ is so chosen that the sum of the areas bounded by $S_n(x)$, the $X$-axis, and the two ordinates, $x = 0$, $x = 1$, is always numerically equal to one. Hence, upon passing to the limit, we have

$$\underset{n\to\infty}{L} \int_0^x S_n(x)dx = x.$$

On the other hand, the series defines a function $f(x)$ which is zero for all values of $x$ in the interval (0, 1) except for a set of points $E$, where $f(x)$ becomes infinite. The set $E$ is of measure zero, since the measure of the set $E_n$ where $S_n(x)$ is different from zero is of measure $\frac{1}{n^2}$, and

$$m(E) = \underset{n\to\infty}{L} \frac{1}{n^2} = 0.$$

The set $E$ is, however, everywhere dense in (0, 1) and consequently $f(x)$ is not integrable.

The given series, therefore, defines a non-integrable function although the series of the indefinite integrals of the individual terms converges to a continuous function.

A condition under which we may integrate a series term by term and obtain the integral of the function defined by the series is given by the following theorem, which is a special case of Theorem III, Art. 46. In this theorem we shall consider the case in which $S_n(x)$ is bounded as to $n$ and $x$ taken together; that is, we shall assume that $S_n(x)$ does not exceed a finite number $M$ for all values of $x$ in the given interval and for all values of $n$ greater than some integer $m$.

THEOREM II. *Given a function $f(x)$ defined in the interval $(a, b)$ by the series $\Sigma\, u_n(x)$, each term of which is an integrable function of $x$. If there exists a finite number $M$ such that for all values of $n$ greater than some integer $m$, and for all values of $x$ in $(a, b)$, we have*

$$|\, S_n(x)\, | < M,$$

*then*

$$\int_a^x f(x)dx = \underset{n\to\infty}{L} \int_a^x S_n(x)dx, \quad a < x \leqq b,$$

*provided the given series converges quasi-uniformly in general in $(a, b)$.*

This theorem is equivalent to saying that if $S_n(x)$ is integrable and bounded as to $n$ and $x$ together and converges to an integrable function $f(x)$, then the given series can be integrated term by term.

Since any series which converges uniformly to a bounded function satisfies the conditions of the foregoing theorem, we have the following corollary.

COR. *Let* $f(x)$ *be a bounded function defined in the interval* $(a, b)$ *by the series* $\Sigma\, u_n(x)$, *where* $u_n(x)$ *is a bounded, integrable function of* $x$ *in* $(a, b)$. *If* $\Sigma\, u_n(x)$ *converges uniformly in* $(a, b)$ *then*

$$\int_a^x f(x)dx = \underset{n\to\infty}{L} \int_a^x S_n(x)dx, \quad a < x \leqq b.$$

It is to be noted that the foregoing condition for the integration of a series term by term applies only to integration over an interval, that is to the definite integral $\int_a^b f(x)dx$ or to the indefinite integral $\int_a^x f(x)dx$, where $a < x \leqq b$. If the integration is taken without reference to an interval but merely as the inverse of differentiation, uniform convergence no longer necessarily gives a sufficient condition for term by term integration of a series, as the following illustration shows.

Given the series

$$\frac{x}{(1+x^2)^2} + \frac{x}{(2+x^2)^2} + \frac{x}{(3+x^2)^2} + \cdot\ \cdot\ \cdot + \frac{x}{(n+x^2)^2} + \cdot\ \cdot\ \cdot\ .$$

This series converges for all positive values of $x$. It is also uniformly convergent, as may be seen from the application of the Weierstrass M-test, since for all positive, finite values of $x$, we have

$$\sum \frac{x}{(n+x^2)^2} \leqq \sum \frac{1}{n^2},$$

and the latter series is convergent.

Integrating each term and assuming in each case the constant of integration to be zero, we have

$$-\frac{1}{2}\left\{\frac{1}{1+x^2} + \frac{1}{2+x^2} + \cdot\ \cdot\ \cdot + \frac{1}{n+x^2} + \cdot\ \cdot\ \cdot\ \right\},$$

which does not converge for the values of $x$ in any finite interval. If, however, we take the indefinite integral $\int_0^x \frac{xdx}{(n+x^2)^2}$, $0 < x \leqq 1$, or the definite integral $\int_0^1 \frac{xdx}{(n+x^2)^2}$ the series of integrals in each case converges and represents the integral, taken between the same limits, of the functions defined by the given series.

It is possible for $S_n(x)$ to be a bounded function of $x$ for each value of $n$ and still not be bounded in $x$ and $n$ together. The series considered in Ex. 3 of Art. 59 furnishes an illustration. The function defined by the given series, while finite, may or may not be bounded in the neighborhood of such a point. As a consequence of Theorem IV, Art. 46, we have the following theorem.

THEOREM III. *Let $f(x)$ be a bounded function, defined in the interval $(a, b)$ by the series $\Sigma u_n(x)$, where each $u_n(x)$ is a bounded, integrable function of $x$. Then if the limit $\underset{n\to\infty}{L}\int_a^x S_n(x)dx$ exists and defines a continuous function, we have*

$$\int_a^x f(x)dx = \underset{n\to\infty}{L}\int_a^x S_n(x)dx, \quad a < x \leqq b,$$

*provided the points of $(a, b)$ in the neighborhood of which $S_n(x)$, considered as a function of $(x, n)$, has no finite upper bound form an enumerable set and the given series converges quasi-uniformly in the given interval.*

If one makes use of the Lebesgue integral, the conditions for term by term integration are somewhat changed. First of all, it is to be observed that if a series of summable functions converges to a bounded function $f(x)$, then this function is summable. This result follows at once from the fact that every summable function is measurable and a sequence of measurable functions defines, by the corollary to Theorem X, Art. 49, a measurable function. If then $f(x)$ is also bounded it is of course summable. From Theorem VII, Art. 51, we have the following theorem for series of measurable functions.

THEOREM IV. *Let $\Sigma\, u_n(x)$ be a series of measurable functions, each of which is bounded as to $(x, n)$ together on the set of points $E$. Then if the series converges to the limiting function $f(x)$, we have*

$$\int_E f(x)dx = \underset{n\to\infty}{L}\int_E S_n(x)dx.$$

In case the terms of the given series are summable we have as a special case of Theorem VI, Art. 53, the following theorem.

THEOREM V. *Let $\Sigma\, u_n(x)$ be a series of summable functions converging to the limiting function $f(x)$ on the set $E$. If we have $|\,S_n(x)\,| < \phi(x)$, $n = 1, 2, 3 \cdots$, where $\phi(x)$ is positive and summable on $E$, then $f(x)$ is summable and*

$$\int_E f(x)dx = \underset{n\to\infty}{L}\int_E S_n(x)dx.$$

Finally, if the terms of the given series $\Sigma\, u_n(x)$ are either zero or positive for all values of $x$ of a given set $E$, we can state a necessary and sufficient condition for the term by term Lebesgue integration of the series as follows [see Theorem VIII, Art. 53].

THEOREM VI. *Let* $\Sigma\, u_n(x)$ *be a series of non-negative, summable functions having the sum* $f(x)$. *Then a necessary and sufficient condition that* $f(x)$ *is summable and that*

$$\int_E f(x)dx = \underset{n\to\infty}{L} \int_E S_n(x)dx,$$

*is that the integrals* $\int S_n(x)dx$ *are equi-absolutely continuous on* $E$.

**62. Differentiation of a series.** One can not differentiate every series term by term and thus obtain the derivative of the function defined by the series. Since the derivative of a finite number of terms is the sum of the derivatives of those terms, it follows that differentiating a series term by term is equivalent to writing

$$\underset{n\to\infty}{L}\, D_x\{S_n(x)\} = D_x\left\{\underset{n\to\infty}{L}\, S_n(x)\right\}.$$

For example, in the series discussed in Ex. 7, Art. 59, we have for $x = 0$

$$\underset{n\to\infty}{L}\, D_x[S_n(x)]_{x=0} = \underset{n\to\infty}{L}\ \underset{\Delta x\to 0}{L} \left\{\frac{1}{\Delta x}\cdot\frac{n\,\Delta x}{1+n^2(\Delta x)^2}\right\} = \underset{n\to\infty}{L}\, n = \infty.$$

On the other hand, we have

$$D_x\left\{\underset{n\to\infty}{L}\, S_n(x)\right\} = D_x\ \underset{n\to\infty}{L}\ \frac{nx}{1+n^2x^2} = D_x\cdot 0 = 0.$$

Consequently, the given series can not be differentiated term by term at $x = 0$.

The following theorem gives a convenient criterion for the term by term differentiation of an infinite series.

THEOREM I. *Given a function* $f(x)$ *defined for the interval* $(a, b)$ *by the convergent series* $\Sigma\, u_n(x)$, *and suppose that the derivative* $u_n'(x)$ *exists for all values of* $n$ *and that the series* $\Sigma\, u_n'(x)$ *converges uniformly in* $(a, b)$. *Then the derivative* $f'(x)$ *exists at* $x_0$, *and*

$$f'(x_0) = \Sigma\, u_n'(x_0)\ ,\ a \leqq x_0 \leqq b.$$

Since the derivative of each term exists, we have

$$S_n'(x) = u_1'(x) + u_2'(x) + \cdot\ \cdot\ \cdot + u_n'(x),$$

for all values of $n$. Since the series $\Sigma\, u_n'(x)$ converges uniformly in $(a, b)$, it follows that for all values of $x$ in this interval we have

$$|\, R_n'(x)\, | < \epsilon,\ n > m, \tag{1}$$

where $R_n(x)$ denotes the remainder of the given series after the first $n$ terms and $m$ is independent of $x$. For convenience, set

$$S\,(n, \Delta x) \equiv \frac{S_n(x + \Delta x) - S_n(x)}{\Delta x}, \tag{2}$$

where $x$ is any point of the interval $(a, b)$. We then have

$$f'(x) = \underset{\Delta x \to 0}{L}\ \underset{n \to \infty}{L}\ S\,(n, \Delta x), \tag{3}$$

$$\Sigma\, u_n'(x) = \underset{n \to \infty}{L}\ \underset{\Delta x \to 0}{L}\ S\,(n, \Delta x). \tag{4}$$

The required relation is then nothing else than the interchange of the order of limits and we can make use of Theorem VI, Art. 24. We shall show that the conditions of this theorem are satisfied.

The first of these conditions is satisfied from the fact that the given series converges; for, we have

$$\underset{n \to \infty}{L}\ S(n, \Delta x) = \frac{f(x + \Delta x) - f(x)}{\Delta x},\ |\,\Delta x\,| < \delta.$$

Moreover, from the existence of $u_n'(x)$, we have

$$\underset{\Delta x \to 0}{L}\ S(n, \Delta x) = S_n'(x).$$

The second of the required conditions is satisfied, because of the convergence of the series of derivatives, thus giving the relation

$$\underset{n \to \infty}{L}\ S_n'(x) = A.$$

To show that the third of the given conditions is also satisfied we shall make use of Cor. 3 to Theorem VII, Art. 31, by which we know that the least upper and greatest lower bounds of a derivative are the same as those of the corresponding difference quotient. Consequently, from (1) we have

$$\left| \frac{R_n(x + \Delta x) - R_n(x)}{\Delta x} \right| < \epsilon,\ \Delta x \neq 0.$$

Remembering that $\qquad R_n(x) = f(x) - S_n(x),$

the foregoing relation may be written in the form

$$\left| \frac{f(x + \Delta x) - f(x)}{\Delta x} - S(n, \Delta x) \right| < \epsilon,$$

for all values of $n > m$ and for all values of $|\,\Delta x\,| < \delta$. This relation, however, is none other than the third condition of the theorem already quoted for the interchange of the order of limits in (3) and (4). Consequently, we have

$$f'(x) = \Sigma\, u_n'(x),$$

where $x$ is any point of the interval $(a, b)$.

This theorem gives a sufficient but not a necessary condition as the following example shows.

**Ex. 1.** Given the series whose first $n$ terms have the sum $S_n(x) = \frac{1}{2n^2}\log(1+n^4x^2)$. Show that the series $\Sigma u_n'(x)$ does not converge uniformly, but that the given series can be differentiated term by term.

We have as the result of term by term differentiation

$$\Sigma u_n'(x) = \underset{n\to\infty}{L}\, D_xS_n(x) = \underset{n\to\infty}{L}\, D_x \frac{1}{2n^2}\log(1+n^4x^2)$$

$$= \underset{n\to\infty}{L}\, \frac{n^2x}{1+n^4x^2} = \underset{n\to\infty}{L}\, \frac{\dfrac{x}{n^2}}{\dfrac{1}{n^4}+x^2} = 0.$$

The given series converges in the interval (0, 1) to the value zero. Consequently, we have for $x$ in this interval

$$f'(x) = 0 = \Sigma u_n'(x).$$

The series of derivatives $\Sigma u_n'(x)$, however, does not converge uniformly, since the simultaneous limit $\underset{\substack{x\to x_0\\ n\to\infty}}{L}\, S_n'(x)$ does not exist at the point $x_0 = 0$.

If we make use of the Lebesgue theory of integration, we can establish the following theorems for term by term differentiation of an infinite series.

THEOREM II. *Given a function $f(x)$ defined for the interval $(a, b)$ by the convergent series $\Sigma\, u_n(x)$, and suppose that the derivative $u_n'(x)$ exists and that the series $\Sigma\, u_n'(x)$ converges to a continuous function $\phi(x)$. Then, if $S_n'(x)$ is bounded as to $x$ and $n$ together, the derivative $f'(x)$ exists and*

$$f'(x) = \Sigma\, u_n'(x).$$

As the series $\Sigma\, u_n'(x)$ converges and defines the function $\phi(x)$, we have

$$\phi(x) = u_1'(x) + u_2'(x) + \cdots + u_n'(x) + \cdots. \tag{5}$$

Since $S_n'(x)$ is bounded as to $x$ and $n$ together, it follows from Theorem VII, Art. 51, that the foregoing series can be integrated term by term in the Lebesgue sense, and hence we have, by use of Theorem V, Art. 54,

$$\int_a^x \phi(x)dx = \int_a^x u_1'(x)dx + \int_a^x u_2'(x)dx + \cdots + \int_a^x u_n'(x)dx + \cdots.$$

$$= u_1(x) - u_1(a) + u_2(x) - u_2(a) + \cdots + u_n(x) - u_n(a) + \cdots.$$

Since the series $\Sigma\, u_n(x)$ converges, the series $\Sigma\, u_n(a)$ must also converge and the foregoing series then converges absolutely. Hence, we may write

$$\int_a^x \phi(x)dx = \Sigma\, u_n(x) - \Sigma\, u_n(a) = f(x) - f(a).$$

Because $\phi(x)$ is continuous, it follows that the indefinite integral $\int_a^x \phi(x)dx$ has $\phi(x)$ as its derivative at each point of $(a, b)$. Consequently $f(x)$ must also have a derivative and we have upon differentiating the foregoing relation

$$\phi(x) = f'(x)$$

as was to be proved.

By aid of Theorem VI, Art. 53, we can state a corresponding theorem for the case where $S_n'(x)$ is non-bounded but less in absolute value than some summable function.

THEOREM III. *Given a function $f(x)$ defined for the interval $(a, b)$ by the convergent series $\Sigma\, u_n(x)$. Suppose that the derivative $u_n'(x)$ not only exists but is bounded and that the series $\Sigma\, u_n'(x)$ converges. If for all values of $n$ the integrals $\int S_n'(x)dx$ are equi-absolutely continuous, then $f'(x)$ exists and*

$$f'(x) = \Sigma\, u_n'(x)$$

*for all values of $x$ in $(a, b)$ with the exception of at most a set of measure zero.*

As we have seen,[1] if the integrals $\int S_n'(x)dx$ are equi-absolutely continuous, then the integrals $\int_a^x S_n(x)dx$ are bounded for all values of $n$. As in the proof of the previous theorem, we let

$$\phi(x) = u_1'(x) + u_2'(x) + \cdot\ \cdot\ \cdot + u_n'(x) + \cdot\ \cdot\ \cdot\ .$$

By hypothesis, the integral of the sum of the first $n$ terms of this series is equi-absolutely continuous and hence summable. By Theorem IV, Art. 53, the function $\phi(x)$ is also summable for values of $x$ in the given interval $(a, b)$. It follows from Theorem V of the same article that we have

$$\int_a^x \phi(x)dx = \underset{n\to\infty}{L} \int_a^x S_n'(x)dx \equiv \int_a^x u_1'(x)dx + \int_a^x u_2'(x)dx + \cdot\ \cdot\ \cdot + \int_a^x u_n'(x)dx + \cdot\ \cdot\ \cdot\ .$$

---

[1] See p. 301.

Since $u_n'(x)$ is bounded for all values of $x$ in $(a, b)$, it follows by Theorem VIII, Art. 38, that $u_n(x)$ is continuous and of limited variation. Consequently, $u_n(x)$ is absolutely continuous,[1] and $u_n'(x)$ satisfies the fundamental law of integration [Theorem III, Art. 54] and we have from the above relation

$$\int_a^x \phi(x)dx = u_1(x) - u_1(a) + u_2(x) - u_2(a) + \cdot \cdot \cdot + u_n(x) - u_n(a) + \cdot \cdot \cdot = f(x) - f(a).$$

Differentiating we have, by Theorem III, Art. 54,

$$\phi(x) = f'(x),$$

for all values of $x$ in $(a, b)$ with the exception of at most a set of points of measure zero.

It may also be shown[2] that in case $f(x)$ is defined by a series $\Sigma\, u_n(x)$ of monotone, increasing functions, then the series can be differentiated term by term for all values of $x$ with the exception of at most a set of points of measure zero.

**63. Power series.** It is not the purpose to include in this chapter a discussion of the properties of various special types of series. However, power series play such an important rôle in the theory of functions that it seems best to depart from this purpose far enough to include some of the general properties of such series. By a **power series** we shall understand a series of the form

$$a_0 + a_1x + a_2x^2 + \cdot \cdot \cdot + a_nx^n + \cdot \cdot \cdot,$$

where $n$ is a positive integer and the $a$'s are constants. For the present purpose we shall also regard the coefficients $a_n$ as real numbers. If $n$ is negative or fractional, the series may be spoken of as power series with negative exponents or with fractional exponents as the case may be.

Among the special properties of power series, we have those given in the following theorems.

Theorem I. *If there exists a value $x_0$ for which all the terms of the power series $\Sigma\, a_nx^n$ are less in absolute value than some number $M$, then the given series converges absolutely for all values of $x$ such that $|x| < |x_0|$.*

---

[1] See remarks following Theorem VII, Art. 38.

[2] See Fubini, *Rend. Atti. Acc. Lincei* (Rome), Vol. 24 (1915), pp. 204–205; also Rajchman and Saks, *Fundamenta Math.*, Vol. 4 (1923), 211–213.

Since, by hypothesis, $|a_n x_0^n| < M$, we have from the given series

$$|a_0| + |a_1 x| + |a_2 x^2| + \cdots + |a_n x^n| + \cdots =$$
$$|a_0| + |a_1 x_0| \left|\frac{x}{x_0}\right| + |a_2 x_0^2| \left|\frac{x}{x_0}\right|^2 + \cdots + |a_n x_0^n| \left|\frac{x}{x_0}\right|^n + \cdots$$
$$\leqq M\left\{1 + \left|\frac{x}{x_0}\right| + \left|\frac{x}{x_0}\right|^2 + \cdots + \left|\frac{x}{x_0}\right|^n + \cdots\right\}.$$

If the series in the second member of this inequality converges, then the given series converges absolutely. However, the series in the second member is a geometric series and converges for those values of $x$ that make each term numerically less than one, that is for those values of $x$ where

$$|x| < |x_0|,$$

which was to be shown.

It is to be noted that a power series is always convergent for $x = 0$.

THEOREM II. *If a given power series $\Sigma\, a_n x^n$ converges for $x = x_0$, then it converges absolutely for those values of $x$ for which $|x| < |x_0|$.*

By hypothesis, the series

$$a_0 + a_1 x_0 + a_2 x_0^2 + \cdots + a_n x_0^n + \cdots$$

converges. It follows that the $n^{\text{th}}$ term must approach zero as $n$ becomes infinite. There must then exist a positive number $M$ such that for all values of $n$ we have

$$|a_n x_0^n| < M,$$

and consequently by Theorem I the given series converges absolutely for $|x| < |x_0|$.

THEOREM III. *If a given power series $\Sigma\, a_n x^n$ is divergent for $x = x_0$, then it diverges for all values of $x$ such that $|x| > |x_0|$.*

Let $x_1$ be any value of $x$ such that $|x_1| > |x_0|$. If $\Sigma\, a_n x^n$ were convergent for $x = x_1$, then by Theorem I it would also converge for $x = x_0$, which is contrary to the given hypothesis. From this contradiction the theorem follows.

A given power series may converge for all values of $x$, or it may converge for some values and diverge for others. Thus the series

$$1 + x + x^2 + x^3 + \cdots + x^n + \cdots$$

converges for all values of $x$ in the open interval $(-1, 1)$ but diverges for all values of $x$ exterior to this interval. It is of importance to determine whether there always exists a point which separates the points of convergence from those of divergence. The following theorem gives the answer to such an inquiry.

THEOREM IV. *For a power series $\Sigma a_n x^n$ which converges for some values of $x \neq 0$ but is not always convergent, there exists a definite number $\rho$ such that for $|x| < \rho$ the series is convergent and for $|x| > \rho$ the series is divergent.*

By hypothesis, there must exist at least one point of convergence different from $x = 0$, and one point of divergence. Denote these points by $a_1$ and $b_1$, respectively. It follows from Theorems II and III that $|a_1| < |b_1|$. Divide the interval $(a_1, b_1)$ into two equal parts. Let $x_0$ be the middle point of this interval. If the given series converges for $x_0$, then we shall further consider the interval $(x_0, b_1)$. If, on the other hand, the series is divergent for $x_0$, we shall consider the interval $(a_1, x_0)$. For the sake of definiteness suppose the interval $(a_1, x_0)$ to be the one to be further subdivided. Denote it by $I_0$. Divide this interval into two equal parts and consider the left half or the right half according as the series diverges or converges at the middle point. Denote the new interval to be considered by $I_1$. Think of this process of subdivision and selection of new intervals to be carried on indefinitely. In this manner we obtain a sequence of intervals $\{I_n\}$ such that by the theorem of Art. 7 they define in the limit a single point, which we shall denote by $c$. For every value of $x$ satisfying the condition that $|x| < c$, the given series is convergent, and for all values of $x$ such that $|x| > c$ the series is divergent. The point $c$ determines the value $\rho$ which the theorem requires.

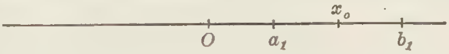


FIG. 49.

Suppose we have a positive value $\rho$ which satisfies the conditions that $\Sigma a_n x^n$ converges for $|x| < \rho$ and diverges for $|x| > \rho$. The interval $(-\rho, \rho)$ is then called the **interval of convergence** of the given power series, and it is often convenient to speak of $\rho$ as the **radius of convergence.** Thus the power series $\sum_{n=0}^{\infty} x^n$ has one as its radius of convergence and its interval of convergence is (-1, 1). The interval of convergence may be an open or a closed interval. To determine the interval of convergence, we have the following theorem.

THEOREM V. *Given the power series $\Sigma a_n x^n$, and denote by $A$ the maximum limiting value of the sequence.*

$$|a_1|, \sqrt{|a_2|}, \sqrt[3]{|a_3|}, \cdot \cdot \cdot, \sqrt[n]{|a_n|}, \cdot \cdot \cdot .$$

*Then the radius of convergence of the series is equal to* $\frac{1}{A}$.

This theorem can be established by comparing the given series with the geometric series $\Sigma a^n x^n$, which converges absolutely if

$$|ax| = |a| \cdot |x| < 1;$$

that is, if

$$|x| < \frac{1}{|a|}.$$

Putting $a^n = a_n$, we see that the given series is convergent, and in fact absolutely, for those values of $x$ for which

$$\sqrt[n]{|a_n x^n|} = \sqrt[n]{|a_n|} \cdot |x| \leqq |ax| < 1.$$

This conclusion is also true if we replace $\sqrt[n]{|a_n|}$ by the maximum limiting value of the sequence $\{\sqrt[n]{|a_n|}\}$; that is, the given series will be absolutely convergent if we have

$$A \cdot |x| \leqq |ax| < 1,$$

or

$$|x| < \frac{1}{A}.$$

Since convergence requires that the absolute value of the terms of the series shall approach zero as $n$ becomes infinite, the given series is divergent if we have for all values of $n$

$$|a_n x^n| \geqq |ax|^n > 1;$$

that is, if

$$\sqrt[n]{|a_n|} \cdot |x| > 1,$$

or

$$|x| > \frac{1}{\sqrt[n]{|a_n|}}.$$

This conclusion is still true if we replace $\sqrt[n]{|a_n|}$ by the maximum limiting value of the sequence of these values as $n$ becomes infinite, that is by $A$. We then have the conclusion that the series $\Sigma a_n x^n$ is divergent whenever

$$|x| > \frac{1}{A}.$$

Since the given series is convergent for $|x|$ less than $\frac{1}{A}$ and divergent for $|x|$ greater than this same value, it follows that $\frac{1}{A}$ is the radius of convergence as the theorem states.

In case the sequence $\sqrt[n]{|a_n|}$ has a definite limit, then of course that limit is itself the radius of convergence.

THEOREM VI. *The power series* $\Sigma a_n x^n$ *converges uniformly and absolutely in any interval which falls entirely within the interval of convergence; that is, for the values of* $x$ *satisfying the condition*

$$|x| \leqq |x_0| < \rho,$$

*where* $\rho$ *is the radius of convergence.*

However close $x_0$ may be to $\rho$, we can always select a value $x_1$ such that $|x_0| < |x_1| < \rho$. The series of absolute values gives

$$|a_0| + |a_1 x| + |a_2 x^2| + \cdot \cdot \cdot + |a_n x^n| + \cdot \cdot \cdot = |a_0| + |a_1 x_1| \frac{|x|}{|x_1|} + |a_2 x_1^2| \frac{|x|^2}{|x_1|^2} + \cdot \cdot \cdot + |a_n x_1^n| \frac{|x|^n}{|x_1|^n} + \cdot \cdot \cdot . \quad (1)$$

By Theorem IV the given series converges for $|x| = |x_1| < \rho$. It follows that

$$\underset{n \to \infty}{L} |a_n x_1^n| = 0,$$

and hence there must exist a definite value $M$ for which

$$|a_n x_1^n| < M, n = 1, 2, 3, \cdot \cdot \cdot .$$

Hence, from (1) we have

$$\Sigma |a_n x^n| < M \left\{ 1 + \frac{|x|}{|x_1|} + \frac{|x|^2}{|x_1|^2} + \cdot \cdot \cdot + \frac{|x|^n}{|x_1|^n} + \cdot \cdot \cdot \right\}.$$

This latter series converges for $|x| < |x_1|$ and therefore for $|x| \leqq |x_0|$. Hence, $\Sigma a_n x^n$ converges for the same values of $x$. However, the value $x_0$ is any value less than the radius of convergence $\rho$. One may, therefore, state the result by saying that the given power series converges absolutely for all values of $x$ in any closed interval lying wholly within the interval of convergence of the given series.

That the series also converges uniformly in the interval for which $|x| \leqq |x_0|$, follows at once from the application of the Weierstrass test for uniform convergence. The largest numerical values of the terms of the given series are obtained by taking the extreme values in the closed interval $(-x_0, x_0)$. As the given series converges absolutely for such values, it follows that the Weierstrass test is satisfied and hence the given series converges uniformly in this interval.

It is to be noted that while the interval of uniform convergence is necessarily closed that of absolute convergence may be closed or open.

Since the terms of a power series $\Sigma a_n x^n$ are continuous functions of $x$, it follows from the foregoing theorem that the function defined by such a series is also continuous. We may, therefore, state the following theorem.

THEOREM VII. *A power series $\Sigma a_n x^n$ defines a continuous function for all values of $x$ in the interval $(-x_0, x_0)$, where $|x_0|$ is less than the radius of convergence.*

At the end-points of the interval of convergence, the continuity of the function defined by a power series is determined by the following theorem.

THEOREM VIII.[1] *If the power series $\Sigma a_n x^n$ converges for $x = \rho$, where $\rho$ is the radius of convergence, then the function $f(x)$ defined by this series is continuous on the left at $x = \rho$; that is, $\underset{x \to \rho}{L} \Sigma a_n x^n = \Sigma a_n \rho^n$. A similar statement holds for the right-hand continuity at $x = -\rho$ if the series converges at that point.*

We shall prove the first part of the theorem. A similar proof applies to the second part.

It is to be noted that we may, without loss of generality, assume $\rho = 1$. For, if instead of the series $\Sigma a_n x^n$ having the radius $\rho$ we take the series $\Sigma\, b_n x^n$, where $b_n = \dfrac{a_n}{\rho^n}$, we have a series whose radius of convergence is one. Assuming, then, that $\rho = 1$ and that the series $\Sigma a_n x^n$ converges for $x = \rho$, we have

$$f(x) = \Sigma a_n x^n, \tag{2}$$

and

$$f(1) = \Sigma a_n = A.$$

Since we have

$$\frac{1}{1-x} = \Sigma x^n, \text{ for } |x| < 1, \tag{3}$$

we may write

$$\frac{1}{1-x}\Sigma a_n x^n = \Sigma x^n \Sigma a_n x^n = \Sigma S_n x^n, \tag{4}$$

where

$$S_n = a_0 + a_1 + a_2 + \cdot\ \cdot\ \cdot + a_n.$$

From (2) and (4) we have

$$f(x) = (1 - x)\Sigma S_n x^n. \tag{5}$$

---

[1] See Abel, *Jour. für die reine und angew. Math.*, Vol. 1 (1826), p. 311.

From (3) we obtain $1 = (1 - x)\Sigma x^n,$

and consequently it follows that

$$A = \Sigma a_n = (1 - x)\Sigma A x^n. \tag{6}$$

Subtracting (5) from (6), we obtain

$$A - f(x) = (1 - x)\,\Sigma(A - S_n)x^n. \tag{7}$$

For $n$ sufficiently large, the difference $|A - S_n|$ is arbitrarily small, and hence we may write

$$|A - S_n| < \frac{\epsilon}{2},\ n > m.$$

From (7), we have for $|x| < 1$

$$|A - f(x)| \leqq |(1 - x)\sum_{n=0}^{m}(A - S_n)| + \frac{\epsilon}{2}\left|(1 - x)\sum_{n=m+1}^{\infty} x^n\right|\cdot \tag{8}$$

Put $|A - S_1| + |A - S_2| + \cdot\cdot\cdot + |A - S_m| = B,$

and we have from (8)

$$|A - f(x)| \leqq B(1 - x) + \frac{\epsilon}{2}\left|(1 - x)\cdot\frac{x^{m+1}}{1 - x}\right|\cdot \tag{9}$$

If we now put $\delta = \dfrac{\epsilon}{2B},$

where $1 - \delta < x < 1$, we get from (9)

$$|A - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which shows that $f(x)$ approaches $A$ as $x$ approaches 1 on the left.

As we have seen, a series whose terms are continuous functions can be integrated term by term in any closed interval in which the series converges uniformly. Moreover, if we differentiate a power series term by term, we obtain a power series, which again converges uniformly by Theorem VI. We may, therefore, state the following theorem.

THEOREM XI. *Every power series may be integrated or differentiated term by term in any closed interval which lies within the interval of convergence of the given series.*

To determine the interval of convergence of the series obtained by integrating or differentiating a power series term by term, we have the following theorem.

THEOREM X. *The series resulting from the differentiation or integration of a power series term by term has the same interval of convergence as the original series.*

Let the given series be $\Sigma a_n x^n$. We are to show that the series $\Sigma n a_n x^{n-1}$ and $\Sigma \dfrac{a_n}{n+1} x^{n+1}$ have the same interval of convergence as the series $\Sigma a_n x^n$.

The radius of convergence of $\Sigma a_n x^n$ is the reciprocal of the maximum limiting value of the sequence $\{\sqrt[n]{|a_n|}\}$. The reciprocal of the radius of convergence of the series $\Sigma n a_n x^{n-1}$ is then

$$\overline{L}_{n\to\infty} \sqrt[n]{|n \cdot a_n|} = \overline{L}_{n\to\infty} \sqrt[n]{|a_n|} \cdot \sqrt[n]{n} = \overline{L}_{n\to\infty} \sqrt[n]{|a_n|},$$

since the limit of $\sqrt[n]{n}$ as $n$ becomes infinite is one. Similarly, for the series $\Sigma \dfrac{a_n}{n+1} x^n$, the reciprocal of the radius of convergence is given by

$$\overline{L}_{n\to\infty} \sqrt[n]{\frac{a_n}{n+1}} = \overline{L}_{n\to\infty} \sqrt[n]{|a_n|} \cdot \sqrt[n]{\frac{1}{n+1}} = \overline{L}_{n\to\infty} \sqrt[n]{|a_n|},$$

because the limit of $\sqrt[n]{\dfrac{1}{n+1}}$ is also one. As $\overline{L}_{n\to\infty} \sqrt[n]{|a_n|}$ is the reciprocal of the radius of convergence of the series $\Sigma a_n x^n$, the theorem is established.

We may not only add and subtract power series, but may multiply such series as the following theorem shows.

THEOREM XI. *Given two power series $\Sigma a_n x^n$, $\Sigma b_n x^n$ converging respectively to the functions $f_1(x)$, $f_2(x)$. We may multiply the series together and thus obtain*

$$f_1(x) \cdot f_2(x) = \Sigma a_n x^n \cdot \Sigma b_n x^n = \Sigma (a_0 b_n + a_1 b_{n-1} + \cdot \cdot \cdot + a_n b_0) x^n$$

*for all values of $x$ within the smaller of the two intervals of convergence.*

This theorem follows as the consequence of Theorem IV, Art. 58, and the fact that power series converge absolutely in any interval smaller than the interval of convergence.

We may obtain the power series representation of a given function by means of the Taylor or Maclaurin expansion of a function, subject to the condition that in any case the remainder approaches zero as $n$ becomes infinite. As we have seen, we may operate with power series in the same manner as with polynomials, including term by term differentiation and integration. Because of these properties, power series play such an important rôle in the development of the theory of functions that those functions which can be represented by power series are given a special name. They are called **analytic functions.**

From the discussion of power series, it follows at once that an analytic function has among other properties the one that it is continuous and has a continuous derivative of any finite order within the interval for which it is defined.

**64. Condensation of Singularities.** As we have seen, the terms of a series may be discontinuous functions and still the series be uniformly convergent. From the definition of uniform convergence, we have for all values of $n$ larger than some $m$ independent of $x$ and for all values of $x$ in a given interval

$$|f(x) - S_n(x)| < \epsilon,$$

where $\epsilon$ is arbitrarily small. This relation permits the points of discontinuity of the individual terms to be also points of discontinuity of the function defined by the series. Hankel[1] has devised a method by which we can introduce a discontinuity of a predetermined character at the set of rational points, that is at a set of points which is everywhere dense. This process is known as the **condensation of singularities.**

For this purpose, consider a function $\phi(y)$ which is bounded and continuous in the interval $(-1, 1)$ except at the origin, where it has a singularity of the required character. If we replace $y$ by $\sin n\pi x$, the result depends upon $n$ as well as $x$, and we often express that fact by writing $\phi_n(\sin n\pi x)$. As $x$ takes all real values, $y$ takes values between $-1$ and $+1$. Consider now the series $\Sigma c_n \phi_n(\sin n\pi x)$, where $c_n$ is so chosen as to insure the uniform convergence of the series. For each rational value of $x$, that is for $x = p/q$, there are an infinite number of terms of the series $\Sigma c_n \phi_n(\sin n\pi x)$ which have the same singularity as was assigned to $\phi(y)$ at $y = 0$; because, for all values of $n = k \cdot q$, $k = 1, 2, 3, \cdot \cdot \cdot$, $n\pi x$ then becomes a multiple of $\pi$ and the function $\sin n\pi x$ is zero. Under certain restrictions to be noted later, each of the points of discontinuity of the terms $c_n\phi_n(\sin n\pi x)$ are also points of discontinuity of the function $f(x)$ defined by the relation

$$f(x) = \Sigma c_n \phi_n(\sin n\pi x).$$

As already pointed out, a singularity which occurs at a rational point is also present in an infinite number of terms of the series $\Sigma c_n\phi_n(\sin n\pi x)$. It must be shown in any particular case that these singularities do not neutralize one another and thus not appear in the limiting function $f(x)$. It is certain, however, that no discontinuity can appear in $f(x)$ which does not occur in some term of the series because of the uniform convergence of the given series. The following illustrations will serve to make clear the use and the restrictions of Hankel's method of condensation of singularities.

---

[1] See Tübinger *Universitätsschriften aus dem Jahre* 1870, No. 3, p. 6.

Making use of Ex. 2, Art. 25, we have a function $\phi(y)$ which is zero for $y = 0$ and takes the value one for all other values of $y$ in the interval $(-1, 1)$. Replacing $y$ by $\sin n\pi x$ and putting $c_n = \frac{1}{n!}$ we define the function $f(x)$ by the relation

$$f(x) = \Sigma \frac{1}{n!} \phi(\sin n\pi x).$$

The first term of this series has a discontinuity equal to one at the points $x = 0$, $x = \pm 1$. The second term has a discontinuity of $\frac{1}{2}$ at the points

$$x = 0, \pm \tfrac{1}{2}, \pm 1.$$

Likewise the third term has a discontinuity of $\frac{1}{6}$ at the points

$$x = 0, \pm \tfrac{1}{3}, \pm \tfrac{2}{3}, \pm 1.$$

The general term, $\frac{1}{n!} \phi(\sin n\pi x)$, has points of discontinuity with a saltus equal to $\frac{1}{n!}$ at

$$x = 0, \pm \frac{1}{n}, \pm \frac{2}{n}, \pm \frac{3}{n}, \cdot \cdot \cdot, \pm \frac{n-1}{n}, \pm 1.$$

The given series converges uniformly; for, the terms are in absolute value not greater than the terms of the convergent series $\Sigma \frac{1}{n!}$ and consequently conditions for the Weierstrass test for uniform convergence are satisfied. Because of the uniform convergence of the series, the points of discontinuity of $f(x)$ must be among those values of $x$ for which one or more of the terms of the given series are discontinuous. However, these points are always rational and hence we can conclude at once that $f(x)$ is continuous for all irrational values of $x$.

We shall now show that $f(x)$ is discontinuous at all the rational values of $x$; that is, at all those values where $x = \frac{p}{q}\cdot$ For this purpose set up the relation

$$f\left(\frac{p}{q} + \delta\right) - f\left(\frac{p}{q}\right) = \sum_{n=1}^{\infty} \frac{\phi\left[\sin n\left(\frac{p}{q} + \delta\right)\pi\right] - \phi\left[\sin n \frac{p}{q}\pi\right)\right]}{n!}.$$

For those values of $n$ which are a multiple of $q$, say $n = n_1q$, the second term of the numerator vanishes and the first term becomes

$$\begin{aligned}\sin(n_1p\pi + n_1q\delta\pi) &= \sin n_1p\pi \cos n_1q\delta\pi + \cos n_1p\pi \sin n_1q\delta\pi \\ &= \pm \sin n_1q\delta\pi,\end{aligned}$$

where the plus sign is taken if $n_1p$ is even and the negative sign if it is odd. We may then write

$$\begin{aligned}f\left(\frac{p}{q} + \delta\right) - f\left(\frac{p}{q}\right) &= \Sigma' \frac{\phi\left[\sin n\pi\left(\frac{p}{q} + \delta\right)\right] - \phi\left[\sin n\pi \frac{p}{q}\right]}{n!} \\ &\quad + \Sigma \frac{\phi(\pm \sin n_1q\delta\pi)}{n_1q!},\end{aligned}$$

where $\Sigma'$ indicates the sum taken by omitting those terms where $n$ is a multiple of $q$. As these series converge uniformly, the limit of the sum is equal to the sum of the limits of the individual terms. As $\delta$ approaches zero, the limit of the terms in the first series is zero, and hence we have

$$\underset{\delta\to 0}{L}\left[f\left(\frac{p}{q} + \delta\right) - f\left(\frac{p}{q}\right)\right] = \Sigma \underset{\delta\to 0}{L} \frac{\phi(\pm \sin n_1q\delta\pi)}{n_1q!}.$$

From the definition of $\phi(y)$, it follows that

$$\phi(\pm\sin n_1q\delta\pi) = 1, \text{ for } \delta \neq 0.$$

Hence, we have $$\underset{\delta\to 0}{L} \phi(\pm\sin n_1q\delta\pi) = 1,$$

and therefore $$\underset{\delta\to 0}{L}\left[f\left(\frac{p}{q} + \delta\right) - f\left(\frac{p}{q}\right)\right] = \Sigma\frac{1}{n_1q!}.$$

Since the series $\Sigma \frac{1}{n_1q!}$ converges and has a sum different from zero, it follows that $f(x)$ is discontinuous for $x = \frac{p}{q}$, that is at all rational points.

It has therefore been shown that $f(x)$ is point-wise discontinuous having the irrational points as points of continuity and at the rational points a discontinuity of the same type as assigned to the function $\phi(y)$ at $y = 0$.

By the use of the condensation of singularities, one may set up a function $f(x, y)$ which is everywhere continuous in $x$ and in $y$ but discontinuous in $(x, y)$ together at all points where both $x$ and $y$ are rational and continuous in $(x, y)$ together at all other points.

To construct such a function, consider the region defined by $0 \leqq x \leqq 1, 0 \leqq y \leqq 1$ and let

$$\begin{aligned} \phi(x, y) &= \frac{xy}{x^2 + y^2}, \text{ for } x, y \text{ not both zero,} \\ &= 0, \text{ for } x = 0,\ y = 0. \end{aligned}$$

We have seen in Art. 28 that this function has an $(x, y)$-discontinuity at the origin. Replace $x$ and $y$ by $\sin^2 n\pi x$ and $\sin^2 n\pi y$, respectively. We have then

$$\phi_n(x, y) = \frac{\sin^2 n\pi x \cdot \sin^2 n\pi y}{\sin^4 n\pi x + \sin^4 n\pi y}.$$

This function is everywhere continuous in $x$ alone and in $y$ alone. We shall now examine it for points of discontinuity in $(x, y)$ together. For $n = 1$, we have

$$\phi_1(x, y) = \frac{\sin^2 \pi x \cdot \sin^2 \pi y}{\sin^4 \pi x + \sin^4 \pi y}.$$

This function has at the points

$$(0, 0),\ (0, 1),\ (1, 0),\ (1, 1)$$

an $(x, y)$-discontinuity, because at each of these points the simultaneous limit in $(x, y)$ does not exist. We shall examine the behavior of $\phi_1(x, y)$ at the point $(1, 0)$. At the other points, it may be examined in a similar manner. If we put $x = y + 1$ and allow $y$ to approach zero, we have

$$\underset{y \to 0}{L} \frac{\sin^2 \pi(y + 1) \cdot \sin^2 \pi y}{\sin^4 \pi(y + 1) + \sin^4 \pi y} = \underset{y \to 0}{L} \frac{\sin^2 \pi y \cdot \sin^2 \pi y}{\sin^4 \pi y + \sin^4 \pi y} = \frac{1}{2}.$$

If, however, we put $x = 2y + 1$, we obtain

$$\underset{y \to 0}{L} \frac{\sin^2 \pi(2y + 1) \cdot \sin^2 \pi y}{\sin^4 \pi(2y + 1) + \sin^4 \pi y} = \underset{y \to 0}{L} \frac{4 \cos^2 \pi y}{16 \cos^4 \pi y + 1} = \frac{4}{17}.$$

As these two limiting values are different, it follows that the simultaneous limit $\underset{\substack{x \to 1 \\ y \to 0}}{L} \phi_1(x, y)$ does not exist and hence $\phi_1(x, y)$ is discontinuous at this point. In a manner similar to that used above, we may show that $\phi_2(x, y)$ has an $(x, y)$-discontinuity at each of the points $(0, 0)$ $(0, \frac{1}{2})$, $(0, 1)$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, $(1, 0)$, $(1, \frac{1}{2})$, $(1, 1)$; that is, at the points where

$$\begin{aligned} x &= 0, \tfrac{1}{2}, 1, \\ y &= 0, \tfrac{1}{2}, 1. \end{aligned}$$

At all other points $\phi_2(x, y)$ is continuous in the two variables together. The general term

$$\phi_n(x, y) = \frac{\sin^2 n\pi x \cdot \sin^2 n\pi y}{\sin^4 n\pi x + \sin^4 n\pi y}$$

is discontinuous in $(x, y)$ at the points where

$$\left.\begin{aligned} x &= 0, \frac{1}{n}, \frac{2}{n}, \cdot\ \cdot\ \cdot, \frac{n}{n}\,(=1) \\ y &= 0, \frac{1}{n}, \frac{2}{n}, \cdot\ \cdot\ \cdot, \frac{n}{n}\,(=1). \end{aligned}\right\} \tag{1}$$

This statement can be established by the method already applied in the discussion of $\phi_1(x, y)$; that is, by the two substitutions

$$x = \ y' + \frac{\alpha}{n},\ y = y' + \frac{\beta}{n},$$

and

$$x = 2y' + \frac{\alpha}{n},\ y = y' + \frac{\beta}{n},$$

where $\frac{\alpha}{n}, \frac{\beta}{n}$ denote any of the values of $x, y$, respectively, given in (1). These substitutions lead to different limiting values as $y'$ approaches zero and hence $\phi_n(x, y)$ is discontinuous in $(x, y)$ together at the points in question.

We shall now define the required function $f(x, y)$ by the relation

$$f(x, y) = \sum_{n \to 1}^{\infty} \frac{1}{n!}\,\phi_n(x, y).$$

This series converges for all values of the variables in the given region; for, since $\phi_n(x, y)$ is always less than or equal to $\frac{1}{2}$, we have

$$\Sigma \frac{1}{n!}\,\phi_n(x, y) \leqq \frac{1}{2}\,\Sigma \frac{1}{n!},$$

and $\Sigma \frac{1}{n!}$ is a well-known convergent series. It follows that $f(x, y)$ is fully defined for the given region.

It remains to show that $f(x, y)$ is continuous in each variable separately and then to examine this function for points of discontinuity in $(x, y)$ together. The series defining $f(x, y)$, considered as a function of $(x, n)$ or of $(y, n)$, satisfies the condition for uniform convergence. As the individual terms are continuous in each variable separately, it follows that $f(x, y)$ is likewise continuous in each variable separately.

We shall now show that $f(x, y)$ has an $(x, y)$-discontinuity at a set of points everywhere dense in the given region. As we have seen, the first term has points of $(x, y)$-discontinuity at the points

$$(0, 0),\ (0, 1),\ (1, 0),\ (1, 1).$$

The second term contributes to the saltus at the above points and introduces in addition new points of discontinuity, namely,

$$(0, \tfrac{1}{2}),\ (\tfrac{1}{2}, 0),\ (\tfrac{1}{2}, \tfrac{1}{2}),\ (\tfrac{1}{2}, 1),\ (1, \tfrac{1}{2}).$$

In general each term has some of the points of discontinuity of the preceding terms and also certain new points of discontinuity. These points of discontinuity cannot neutralize one another, since $\phi_n(x, y)$ is always zero or positive and hence those terms which do not vanish are always positive and the amount of the saltus in $(x, y)$ at any point is increased rather than diminished by the addition of new terms of the series. We may choose $n$ sufficiently large so that any particular rational point becomes a point of discontinuity; that is, any point where

$$x = \frac{p}{q},\ y = \frac{p'}{q'},$$

$p$, $p'$, $q$, $q'$ being all integers and $q$, $q'$ different from zero and prime respectively to $p$, $p'$. For this purpose we need only to take $n$ sufficiently large to be divisible by $q$ and $q'$. When this is the case, then $n\pi x$ and $n\pi y$ are integral multiples of $\pi$ and $\phi_n(x, y)$ contains only the sines of integral multiples of $\pi$. The amount of the discontinuity at the various rational points is not the same, because in every case the saltus of $\phi_n(x, y)$ is divided by $n!$ and the same terms of the given series do not always contribute to the saltus. In fact, the amount of discontinuity of $f(x, y)$ is always finite and at many rational points is very small, but the essential thing in our discussion is that every such point is a point of $(x, y)$-discontinuity. As these points are everywhere dense in the given region, the points of $(x, y)$-discontinuity of $f(x, y)$ are likewise everywhere dense. That the points of continuity of $f(x, y)$ are also everywhere dense follows from the fact that $f(x, y)$ can be at most point-wise discontinuous.

As a further illustration of the Hankel method of condensation of singularities, we shall set up a function $f(x, y)$ which is totally differentiable at a set of points everywhere dense and yet is such that there exists also a set of points everywhere dense in the given region where $f(x, y)$ is not totally differentiable. In the region $|x| \leqq 1$, $|y| \leqq 1$, suppose we consider the function

$$\phi(x, y) = \frac{x^2 y}{x^4 + y^2},$$

if $x$, $y$ are not both zero, and where $\phi(0, 0) = 0$. This function is not continuous in $x$ and $y$ together at the origin; for, we obtain different limiting values by taking different approaches to that point. As continuity of $\phi(x, y)$ in $x$ and $y$ together is a necessary condition for total differentiability, it follows that the given function is not totally differentiable at $x = 0$, $y = 0$.

For $x \neq 0$, $y \neq 0$, we have

$$\phi'_x = 2xy \cdot \frac{y^2 - x^4}{(y^2 + x^4)^2}.$$

For $(x = 0, y = 0)$, for $(x \neq 0, y = 0)$, and for $(x = 0, y \neq 0)$, we have $\phi'_x = 0$. For $(x \neq 0, y \neq 0)$, we have

$$\phi'_y = x^2 \frac{x^4 - y^2}{(x^4 + y^2)^2};$$

while for $(x = 0,\ y = 0)$, $(x = 0,\ y \neq 0)$, we have $\phi'_y = 0$. For $(x \neq 0, y = 0)$, we obtain $\phi'_y = \frac{1}{x^2}$.

Although the given function is not totally differentiable at the origin, yet in every neighborhood of that point including the origin itself, the two partial derivatives $\phi'_x, \phi'_y$ exist and one of them, namely $\phi'_x$, is continuous in $x$ alone and in $y$ alone. The partial derivative $\phi'_y$, is everywhere continuous with respect to $y$. It is also continuous with respect to $x$, except at $x = 0$, $y = 0$.

As we have already seen,[1] if the partial derivatives $f'_x, f'_y$ exist for all values of $(x, y)$ in the given region, then the points at which $f(x, y)$ is totally differentiable form a set everywhere dense in that region. However, the points at which the given function $f(x, y)$ is not totally differentiable may also form a set of points everywhere dense in the same region, as we shall now show.

Consider the function

$$f(x, y) = \sum_{n=1}^{\infty} \frac{1}{n!} \phi_n(x, y), \tag{2}$$

where $\phi_n(x, y)$ is formed from the function $\phi(x, y)$ considered above by replacing $x$ by $\sin^2 n\pi x$ and $y$ by $\sin^2 n\pi y$; that is,

$$\phi_n(x, y) = \frac{\sin^4 n\pi x \cdot \sin^2 n\pi y}{\sin^8 n\pi x + \sin^4 n\pi y}.$$

---

[1] See Theorem IV, Art. 34.

The function $\phi(x, y)$ takes the value one-half for $y = x^2 \neq 0$, but for all other values of $x$ and $y$ in the given region it is less than one-half. The amount of the $(x, y)$-discontinuity at the origin is one. Since the sine is never greater than unity, we have

$$\sum_{n=1}^{\infty} \frac{1}{n!} \phi_n(x, y) < \sum_{n=1}^{\infty} \frac{1}{n!}.$$

As the latter series converges, the series $\Sigma \frac{1}{n!} \phi_n(x, y)$ converges uniformly as a function of the two variables $(x, y)$ and hence as a function of either variable separately. As we shall see, the function $f(x, y)$ is not continuous in $x$ and $y$ together throughout the region of definition, because each term of the series is discontinuous at certain points in these two variables. However, $\phi_n(x, y)$ is continuous in $x$ alone and in $y$ alone; and because of the uniform convergence of the above series, $f(x, y)$ is continuous in each variable separately. Consequently, the points at which it is continuous in both variables must form a set everywhere dense.

We shall now show that the points at which $f(x, y)$ has a discontinuity in $x$ and $y$ together also form a set everywhere dense, being those points where both $x$ and $y$ have rational values.

The first term of the series (2) has a discontinuity in $x$ and $y$ at

$$(0, 0),\ (0, \pm 1),\ (\pm 1, 0),\ (\pm 1, \pm 1);$$

the second term at the points

$$|x| = 0, \tfrac{1}{2}, 1,\quad |y| = 0, \tfrac{1}{2}, 1.$$

In the general term we have such a singularity at points $(x, y)$ for which we have any combination of the following values:

$$|x| = 0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ . \ . \ . \ , \frac{n-1}{n}, 1,$$

$$|y| = 0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ . \ . \ . \ , \frac{n-1}{n}, 1.$$

It follows that any point whose coördinates $(x, y)$ are both rational numbers is a point of discontinuity of some term in the series defining $f(x, y)$. It will be observed that none of the terms of the series given in (2) are ever negative, whatever values may be assigned to $x$ and $y$. The points of discontinuity of any term are likewise points of discontinuity of an infinite number of subsequent terms but not necessarily of previous terms. Moreover, the discontinuities at a given point can not be combined so as to cancel one another or to cause the function $f(x, y)$ to have an infinite discontinuity at any point. For, if a point first appears as a discontinuity, say in the $k$th term, the sum of the discontinuities of subsequent terms at that point can not equal in amount the discontinuity of the $k$th term. For example, consider the discontinuity which appears at the rational point $(x = \frac{2}{3}, y = \frac{1}{3})$. This point will appear as a discontinuity for the first time in the third term. The amount of the discontinuity in this term is $\frac{1}{2}\cdot\frac{1}{3!} = \frac{1}{2}\cdot\frac{1}{6}\cdot$ The amount of the discontinuity of subsequent terms at this point can not exceed

$$\frac{1}{4!}\cdot\frac{1}{2} + \frac{1}{5!}\cdot\frac{1}{2} + \cdot\cdot\cdot < \frac{1}{2}\cdot\frac{1}{4!}\left(1 + \frac{1}{5} + \frac{1}{5^2} + \frac{1}{5^3} + \cdot\cdot\cdot\right) = \frac{1}{2}\cdot\frac{1}{4!}\cdot\frac{5}{4} < \frac{1}{2}\cdot\frac{1}{6}.$$

Consequently, all points which appear as points of discontinuity of any term are also points of discontinuity of $f(x, y)$, and hence any point whose coördinates are both rational numbers is such a point. A function must be continuous in the two variables together in order to be totally differentiable. It follows then that $f(x, y)$ as defined is not totally differentiable at any point where the two coördinates are both rational numbers. Such a set of points is everywhere dense in the given region.

We shall now show that the points at which $f(x, y)$ is totally differentiable are also everywhere dense in the given region. To do this we proceed as follows. As we have seen, $\phi'_x$ is continuous in $x$ in the closed interval $-1 \leqq x \leqq 1$; hence for any constant value of $y$, say $y = y_0$, the numerical values of $\phi'_x(x, y_0)$ have a positive finite upper bound. This upper bound, which we shall denote by $M_0$, may change with the choice of $y_0$, but the essential thing is that for each value of $y_0$ it is finite. For $y = y_0$, we have, since $\sin n\pi x$ and $\cos n\pi x$ are numerically never greater than one,

$$\left|\frac{\partial \phi_n}{\partial x}\right| \leqq 2\pi n M_0 \mid \sin n\pi x \cdot \cos n\pi x \mid \leqq 2\pi n \cdot M_0.$$

Hence, we have for a constant value of $y$

$$\sum_{n=1}^{\infty}\left|\frac{1}{n!}\,\frac{\partial \phi_n}{\partial x}\right| \leqq \sum_{n=1}^{\infty}\frac{2\pi M_0}{(n-1)!}.$$

As this last series converges for $-1 \leqq x \leqq 1$, it follows that the series $\Sigma\frac{1}{n!}\cdot\frac{\partial \phi_n}{\partial x}$ converges uniformly in this same interval, and we may write

$$\frac{\partial f}{\partial x} = \sum_{n=1}^{\infty}\frac{1}{n!}\,\frac{\partial \phi_n}{\partial x}.$$

Moreover, since $\frac{\partial \phi_n}{\partial x}$ is continuous in $x$, it follows that $\frac{\partial f}{\partial x}$ is defined by a uniformly convergent series of continuous functions and is therefore itself continuous in $x$.

By a similar method, it may be shown that $\frac{\partial f}{\partial x}$ is continuous in $y$, and since this derivative is also continuous in $x$, it follows that there exists a set of points everywhere dense where it is continuous in $x$ and $y$ together.

We shall now consider the existence of the partial derivative $\frac{\partial f}{\partial y}$. From the discussion of $\phi(x, y)$, it follows that $\phi_y'$ is continuous in $y$ in the closed interval $-1 \leqq y \leqq 1$. Then, for any constant value of $x$ it is bounded, and hence we may write

$$\left| \frac{\partial \phi_n}{\partial y} \right| \leqq 2n\pi M_1 \mid \sin n\pi y \cdot \cos n\pi y \mid \leqq 2n\pi M_1,$$

where $M_1$ is a positive constant for each previously selected value of $x$. Consequently, we have

$$\sum_{n=1}^{\infty} \left| \frac{1}{n!} \frac{\partial \phi_n}{\partial y} \right| \leqq \sum_{n=1}^{\infty} \frac{2\pi M_1}{(n-1)!}.$$

Since this last series converges, it follows that $\Sigma \frac{1}{n!} \frac{\partial \phi_n}{\partial y}$ converges uniformly, and hence we may write

$$\frac{\partial f}{\partial y} = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial \phi_n}{\partial y}.$$

Therefore, the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ both exist at every point in the given region and $\frac{\partial f}{\partial x}$ is continuous in $x$ and $y$ together at a set of points everywhere dense. It follows that $f(x, y)$ is totally differentiable at each of these points.

We have then a function which is totally differentiable at a set of points everywhere dense, and at the same time there exists a set of points likewise everywhere dense at which this function is not totally differentiable, in spite of the fact that both of the first partial derivatives exist at every point and one of them is everywhere continuous in each variable separately. Geometrically, the corresponding surface has a tangent plane which is not parallel to the $Z$-axis at a set of points everywhere dense and there exists simultaneously another set of points at which no such tangent plane exists. It can be shown, moreover, that the points at which these tangent planes exist; that is, the points at which the given function is totally differentiable, must form a set having the cardinal number of the continuum. It will be observed that $\frac{\partial f}{\partial y}$ is not continuous with respect to $x$. By the method used in discussing the continuity of $f(x, y)$ it can be shown that $\frac{\partial f}{\partial y}$ is discontinuous in $x$ at a set of points everywhere dense, namely at the points where $x$ and $y$ are rational numbers.

The foregoing method of condensation of singularities has certain disadvantages. By that method we are able to introduce a given singularity only at the rational points, whereas it is often desirable to select arbitrarily the points at which the singularity is to appear. Moreover, replacing in $\phi(y)$ the variable $y$ by $\sin n\pi x$ introduces some complication due to the fact that the number of the oscillations of the sine function increases indefinitely as $x$ becomes infinite.

Another method, known as the **Cantor method**[1] of condensation of singularities, makes it possible to introduce the given singularity at an arbitrarily chosen enumerable set of points. Let the given set of points be denoted by

$$x_1,\ x_2,\ x_3,\ \cdot\ \cdot\ \cdot,\ x_n,\ \cdot\ \cdot\ \cdot .$$

Suppose the required function $f(x)$ to be defined by the relation

$$f(x) = \sum_{n=0}^{\infty} c_n \phi_n(x - x_n),$$

where the coefficients $c_n$ are positive and so chosen as to insure the uniform convergence of the series in the interval under consideration and, as in the previous discussion, $\phi(y)$ has the value zero for $y = 0$, and aside from this value is a continuous function.

As an illustration, we shall show how we may set up a function $f(x)$ having a derivative at every point in the given interval $(0, 1)$ and this derivative be discontinuous at the set of points $\{x_n\}$. Let

$$f(x) = \Sigma \frac{1}{n^2} \phi_n(x, x_n),$$

where
$$\begin{aligned} \phi_n(x, x_n) &= (x - x_n)^2 \sin \frac{1}{x - x_n}, \text{ for } x \neq x_n \\ &= 0, \text{ for } x = x_n. \end{aligned}$$

The terms of the given series are continuous and the series converges uniformly in the given interval; for, each term is numerically not greater than $\frac{1}{n^2}$ and the series $\Sigma \frac{1}{n^2}$ converges. It follows that $f(x)$ is a continuous function.

---

[1] Although this method was first made use of by Weierstrass, Geo. Cantor was the first to present it in print. See *Math. Ann.* Vol. 19 (1882), p. 588.

By differentiation, we have for $x \neq x_n$

$$\phi_n'(x, x_n) = 2(x - x_n) \sin \frac{1}{x - x_n} - \cos \frac{1}{x - x_n}.$$

For $x = x_n$, we have from the definition of a derivative

$$\phi_n'(x, x_n) = \underset{\Delta x \to 0}{L} \frac{\phi_n(x + \Delta x, x_n) - \phi_n(x, x_n)}{\Delta x} = \underset{\Delta x \to 0}{L} \frac{\Delta^2 x \sin \frac{1}{\Delta x}}{\Delta x} = 0.$$

The series of derivatives, namely

$$\phi_1'(x, x_n) + \frac{1}{2^2} \phi_2'(x, x_n) + \cdot \cdot \cdot + \frac{1}{n^2} \phi_n'(x, x_n) + \cdot \cdot \cdot ,$$

converges uniformly in the given interval (0, 1); because, for all values of $n$ and $x$ we have

$$\phi_n'(x, x_n) < 3,$$

and hence
$$\Sigma \frac{1}{n^2} \phi_n'(x, x_n) < \Sigma \frac{3}{n^2},$$

which shows that the Weierstrass M-test for uniform convergence is satisfied. We can then differentiate the orignal series term by term and write

$$f'(x) = \Sigma \frac{1}{n^2} \phi_n'(x, x_n).$$

Because of the uniform convergence of this series we may also write

$$\underset{x \to x_n}{L} f'(x) = \Sigma \frac{1}{n^2} \underset{x \to x_n}{L} \phi_n'(x, x_n).$$

The limit of the terms expressed in the second member of this relation does not exist because of the non-existence of the limit

$$\underset{x \to x_n}{L} \cos \frac{1}{x - x_n}.$$

Consequently, the limit $\underset{x \to x_n}{L} f'(x)$ can not exist; that is, $f'(x)$ can not be continuous at $x = x_n$. However, the derivative $f'(x)$ exists at each of the points $x = x_n$. Hence, $f(x)$ has a derivative at each point of the given interval (0, 1) but this derivative has at the enumerable set of points $\{x_n\}$ the same type of discontinuity that $\phi'(x)$ has at the origin, when we define $\phi(x)$ by the relation:

$$\phi(x) = x^2 \sin \frac{1}{x}, \text{ for } x \neq 0,$$
$$= 0 \text{ for } x = 0.$$

**65. Divergent Series.** Previous to the nineteenth century, but little attention was paid to the convergence or divergence of the series used in mathematical investigations. Early in that century Fourier and Bolzano set up definitions of convergence. Cauchy formulated more exactly the conditions for convergence, and both he and Abel emphasized the importance of convergence in connection with use of series. In spite of their warning, however, certain well-known divergent series continued to be employed in applied mathematics with satisfactory results. One can see why this was possible; for, a series which converges for a considerable number of terms but ultimately becomes divergent is a divergent series from the point of view of the mathematician, while the physicist or astronomer might find such a series very satisfactory in deducing an empirical formula which would yield results with an error well within the error of observed data. While Stieltjes considered certain important divergent series in attempting a closer approximation to a given function,[1] and Cauchy in 1843 considered the use of Sterling's series for values of the variable for which it is divergent, no real progress was made in the investigation of the conditions under which divergent series may be used until Poincairé considered such a series in connection with his researches on linear differential equations.[2] More recently an extensive literature has appeared, some of the more important results of which will be summarized in this article. The consideration of divergent series may be divided into two parts, the one relating to the theory of asymptotic series and the other to the theory of summability.

A power series derived from a given function by Taylor's expansion does not necessarily represent that function even in the interval of convergence of the resulting series. For example, it is well known that the function

$$f(x) = e^{-\frac{1}{x^2}}$$

gives rise to a power series which converges for all values of $x$ but fails to represent the given function for values of $x \neq 0$. However, for very small values of $x$ the function $\phi(x)$ defined by the series approximates the values of $f(x)$ to a degree depending upon the choice of $x$.

Let us now suppose that, in the interval $(0, b)$, the function $f(x)$ gives rise to the power series

$$a_0 + a_1x + a_2x^2 + \cdot \cdot \cdot + a_nx^n + \cdot \cdot \cdot, \tag{2}$$

---

[1] See *Ann. de l'école normal* (3), Vol. 3 (1866), p. 201.

[2] See *Acta Math.*, Vol. 1 (1886), p. 295.

where we make no assumptions as to the convergence or divergence of this series. Let it be assumed also that

$$\underset{x \to 0+}{L} f(x) = a_0.$$

Likewise, we shall assume that

$$\begin{aligned}
&\underset{x \to 0+}{L} \left\{ \frac{f(x) - a_0}{x} \right\} = a_1, \\
&\underset{x \to 0+}{L} \left\{ \frac{f(x) - a_0 - a_1 x}{x^2} \right\} = a_2, \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&\underset{x \to 0+}{L} \left\{ \frac{f(x) - a_0 - a_1 x - a_2 x^2 - \cdot \cdot \cdot - a_{n-1} x^{n-1}}{x^n} \right\} = a_n, \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot
\end{aligned}$$

If the series (2) is divergent, we can no longer speak of the *sum* of the series in the ordinary sense, but we can say that for a finite number of terms the series gives an approximate value of $f(x)$ for small values of $x$. It will be noted that the smaller $x$ is taken, the larger $n$ must be chosen in order to get the best approximation to the corresponding value of $x$. A series satisfying the foregoing conditions; namely,

$$\underset{x \to 0+}{L} \frac{1}{x^n} \left\{ f(x) - \sum_{n=1}^{n-1} a_n x^n \right\} = a_n, \; n = 0, 1, 2, \cdot \cdot \cdot ,$$

is called an **asymptotic series** and is said to represent the given function $f(x)$ asymptotically for small positive values of $x$.

For large values of $x$, Poincaré introduced a corresponding definition of an asymptotic series.[1] According to this definition, $f(x)$ is said to be represented asymptotically for large positive values of $x$ by the series

$$a_0 + a_1 \left( \frac{1}{x} \right) + a_2 \left( \frac{1}{x} \right)^2 + \cdot \cdot \cdot + a_n \left( \frac{1}{x} \right)^n + \cdot \cdot \cdot , \tag{3}$$

if the following condition is satisfied; namely, if we have for each value of $n$

$$\underset{x \to +\infty}{L} x^n \left\{ f(x) - a_0 - a_1 \left( \frac{1}{x} \right) - \cdot \cdot \cdot - a_{n-1} \left( \frac{1}{x} \right)^{n-1} \right\} = a_n,$$

or what is the same thing, if we have

$$\underset{x \to +\infty}{L} x^n \left\{ f(x) - a_0 - a_1 \left( \frac{1}{x} \right) - \cdot \cdot \cdot - a_n \left( \frac{1}{x} \right)^n \right\} = 0.$$

---

[1] See *Acta Math.*, Vol. 8 (1886), p. 296; see also Ford's *Studies of Divergent Series and Summability* (1916), p. 22.

The relation of $f(x)$ to an asymptotic series representing it is expressed symbolically by writing

$$f(x) \sim a_0 + a_1\left(\frac{1}{x}\right) + a_2\left(\frac{1}{x}\right)^2 + \cdot \cdot \cdot + a_n\left(\frac{1}{x}\right)^n + \cdot \cdot \cdot . \tag{4}$$

The resulting series may be convergent or divergent for the large positive values of $x$ in question. In the case of a convergent series the actual value of $f(x)$ is obtained when $n$ is allowed to become infinite. The best approximation of $f(x)$ by an asymptotic series is obtained if the series is terminated with that term having the smallest absolute value. If the series is of the form $\Sigma\, a_n x^n$, then, as already pointed out, the number of terms necessary increases as $x$ takes smaller values. However, if the series is of the form $\Sigma\, \frac{a_n}{x^n}$, then in order to get the best approximation the number of terms increases as $x$ takes larger values.

Various mathematical processes may give rise to an asymptotic expansion of a function. Frequently, the series arise through the expansion of the given function by means of a Taylor's series. Other methods such as repeated integration by parts may give rise to such an expansion, as the following illustration shows.

Given the function

$$f(x) = \int_x^\infty \frac{e^{x-t}dt}{t}.$$

By repeated integration by parts, we obtain the series

$$\frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \cdot \cdot \cdot + (-1)^{n-1}\frac{(n-1)!}{x^n} + \cdot \cdot \cdot .$$

This series is divergent for all finite values of $x$; for, we have

$$\underset{n\to\infty}{L}\left|\frac{u_n}{u_{n-1}}\right| = \underset{n\to\infty}{L}\left|\frac{(n-1)!}{x^n}\cdot\frac{x^{n-1}}{(n-2)!}\right| = \underset{n\to\infty}{L}\left|\frac{n-1}{x}\right|,$$

which becomes infinite for all finite values of $x$. The above series, however, represents the given function asymptotically for large positive values of $x$. For, we have for the remainder $R_n(x)$ after the first $n$ terms,

$$R_n(x) = (-1)^{n+1}(n+1)!\int_x^\infty \frac{e^{x-t}}{t^{n+2}}dt.$$

As this integral is taken over the interval $(x, \infty)$, we can assume $t \geqq x$ and hence have

$$|\,R_n(x)\,| = (n+1)!\int_x^\infty \frac{e^{x-t}}{t^{n+2}}\,dt < (n+1)!\int_x^\infty \frac{dt}{t^{n+2}} = \frac{n!}{x^{n+1}}.$$

From the definition of asymptotic expansion, we have

$$\underset{x\to+\infty}{L} x^n\{f(x) - S_n(x)\} = \underset{x\to+\infty}{L} \frac{n!}{x} = 0.$$

We may then write for large positive values of $x$

$$f(x) \sim \frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \cdot \cdot \cdot + (-1)^{n-1}\frac{(n-1)!}{x^n} + \cdot \cdot \cdot .$$

From the foregoing discussion, it will be seen that the advantage of an asymptotic series for numerical calculations arises only when the given series happens to converge rapidly for the first few terms so as to bring the error by its use well within a desired limit. It may be shown that the asymptotic representation of a given function is unique; that is, a given function $f(x)$ can have but one such expansion. The converse of this statement, however, is not true; that is, a given asymptotic expansion may represent more than one function.[1]

Two asymptotic expansions

$$\Sigma \frac{a_n}{x^n}, \ \Sigma \frac{b_n}{x^n}$$

are said to be equal if we have $a_n = b_n$ for all values of $n$. It follows from the definition of asymptotic expansions that we can add, subtract, and multiply such expansions and thus obtain an asymptotic expansion of the sum, difference, and product of the given functions just as was done with convergent series. An asymptotic series can also be integrated term by term, thus obtaining an asymptotic expansion of the integral of the given function, provided the first two terms of the expansion are zero and the integral is taken over the interval $(x, \infty)$. One can not, however, always differentiate an asymptotic series term by term and thus obtain an asymptotic expansion of the derivative function without having independent knowledge that the derivative function has an asymptotic expansion in terms of integrable functions different from zero. It is not surprising to find that all asymptotic series can not be differentiated term by term; for, this property is not possessed by all convergent series.

Another class of divergent series have been studied which differ from asymptotic series in that instead of first converging and then diverging they may continuously diverge. A step in this direction was made by Frobenius when he showed[1] that for a power series $\Sigma a_n x^n$ which converges for $x < 1$ we have

---

[1] For a full discussion of the properties of asymptotic series, including formal proofs of those conclusions stated in this article, see Bromwich's *Theory of Infinite Series*, pp. 322–346; also Ford's *Studies on Divergent Series and Summability*, pp. 22–74.

$$\underset{x \to 1}{L} \Sigma\, a_n x^n = \underset{n \to \infty}{L} \frac{S_0 + S_1 + S_2 + \cdot \cdot \cdot + S_n}{n + 1}, \tag{5}$$

where $$S_n = a_0 + a_1 + a_2 + a_3 + \cdot \cdot \cdot + a_n.$$

This result associates with the given series a definite value for $x = 1$ although it is divergent for that value.

Thus, suppose we have

$$f(x) \equiv \frac{1}{1 + x} = 1 - x + x^2 - x^3 + x^4 - \cdot \cdot \cdot \cdot .$$

This series does not converge for $x = 1$, but by the foregoing method we can associate with the series

$$1 - 1 + 1 - 1 + 1 - \cdot \cdot \cdot$$

the result $\frac{1}{2}$. For, we have

$$S_0 = 1,\ S_1 = 0,\ S_2 = 1,\ \cdot \cdot \cdot,\ S_n = \frac{1 + (-1)^n}{2},$$

and hence we have

$$\underset{n \to \infty}{L} \frac{S_0 + S_1 + \cdot \cdot \cdot + S_n}{n + 1} = \underset{n \to \infty}{L} \frac{(n + 1) + \frac{1}{2}[1 + (-1)^n]}{2(n + 1)} = \frac{1}{2}.$$

The value thus obtained is precisely that which we obtain when we give to $x$ the value one in the function which gave rise to the series.

The question to be considered is whether one can so generalize the definition of a sum of a series that under certain conditions we can associate with a divergent series a definite number as its sum which may be used in mathematical discussions. If this can be done by some definite, previously determined plan, then the series is said to be **summable** by this method. Various definitions of summability are to be found in the literature. Any such definition, to be satisfactory, must be such that the same method gives the sum in the ordinary sense when applied for those values of $x$ for which the series is convergent. This property is called the **consistency or regularity** of the definition.[2] The method involved in the definition should also include the **additive property**; that is, if $\sum\limits_{n=0}^{\infty} a_n$ is summable, then $\sum\limits_{n=1}^{\infty} a_n$ is also summable and if the two sums are respectively $S$ and $S'$, then we have

---

[1] See *Journ. für Math.*, Vol. 89 (1880), p. 262.

[2] The term **permancy** is also employed, see Knopp, *Unendliche Reihen*, p. 445.

$$S = a_0 + S'.$$

Moreover, the method should be such that any two summable series can be added or subtracted term by term just as with convergent series.

Two definitions are said to be **mutually consistent** if, whenever each of the definitions gives a value to the series, the two values are the same.

Two definitions are **equivalent** if they have the same range and yield the same sum; that is, if either can be evaluated for a value $x = x_0$, the other can be evaluated for the same value, and the sum is the same by both methods.

It is beyond the scope of this discussion to give a detailed account of the various methods of summability which are to be found in the literature,[1] but a brief summary of one or two of the methods most frequently used will be given.

**Hölder**[2] calls a given series $\Sigma a_n$ summable of order one, if the limit

$$\underset{n\to\infty}{L} \frac{S_0 + S_1 + \cdot \cdot \cdot + S_n}{n+1}$$

exists, where

$$S_n = a_0 + a_1 + \cdot \cdot \cdot + a_n.$$

If this limit does not exist, then he puts

$$S_n^{(1)} \equiv \frac{S_0 + S_1 + \cdot \cdot \cdot + S_n}{n+1}, \quad n = 0, 1, 2, 3, \cdot \cdot \cdot$$

and considers the limit

$$\underset{n\to\infty}{L} \frac{S_0^{(1)} + S_1^{(1)} + \cdot \cdot \cdot + S_n^{(1)}}{n+1}.$$

If this limit exists and is equal to $S$, this value is called the **sum** of the series, and the series is said to be summable of order two. If this limit does not exist then he puts

$$S_n^{(2)} \equiv \frac{S_0^{(1)} + S_1^{(1)} + \cdot \cdot \cdot + S_n^{(1)}}{n+1}$$

---

[1] For a citation and synopsis of the literature on summability, the reader is referred to Smail's *History and Synopsis of the Theory of Summable Infinite Processes.*

[2] See *Math. Ann.*, Vol. 20 (1882), p. 535.

and considers the limit

$$\underset{n\to\infty}{L} \frac{S_0^{(2)} + S_1^{(2)} + \cdots + S_n^{(2)}}{n+1}.$$

In general, the given series is said to be summable of order $r$, if $r$ is the smallest value for which the limit

$$\underset{n\to\infty}{L} S_n^{(r)} \equiv \underset{n\to\infty}{L} \frac{S_0^{(r-1)} + S_1^{(r-1)} + \cdots + S_n^{(r-1)}}{n+1}$$

exists.

**Cesàro**[1] starts with the same arithmetical mean as Hölder and says the series $\Sigma\, a_n$ is summable with simple indeterminancy if the limit

$$\underset{n\to\infty}{L} \frac{S_0 + S_1 + \cdots + S_n}{n+1}$$

exists. He then generalizes the notion of indeterminancy, or order, somewhat differently from Hölder in that he says that the given series $\Sigma a_n$ is summable of order $r$, if $r$ is the smallest positive integer for which the limit

$$\underset{n\to\infty}{L} \frac{S_n^{(r)}}{D_n^{(r)}},\ r > 0, \tag{6}$$

exists, where

$$S_n^{(r)} \equiv S_n + rS_{n-1} + \frac{r(r+1)}{2!}S_{n-2} + \cdots + \frac{r(r+1)\cdots(r+n-1)}{n!}S_0,$$

$$D_n^{(r)} \equiv \frac{(r+1)(r+2)\cdots(r+n)}{n!}.$$

It is of interest to note that the methods of Hölder and Cesàro are equivalent.[2]

**Ex. 1.** Consider the summability by Cesàro's method of the series

$$1 - 2 + 3 - 4 + 5 - 6 + \cdots.$$

We have $S_0 = 1,\ S_1 = -1,\ S_2 = 2,\ S_3 = -2, \cdots.$

For $r = 1$, we have from the above definition

$$S_n^{(1)} = 0, \text{ if } n \text{ is odd,}$$

$$= \frac{n+2}{2(n+1)}, \text{ if } n \text{ is even.}$$

$$D_n^{(1)} = n + 1.$$

Consequently, the limit $\underset{n\to\infty}{L} \frac{S_n^{(1)}}{D_n^{(1)}}$ does not exist.

---

[1] See *Bulletin des Sci. Math.* (2), Vol. 14 (1890), p. 114.

[2] See Schnee, *Math. Ann.*, Vol. 67 (1908), pp. 110–125; also Schur, *Math. Ann.* Vol. 74 (1913), pp. 447–458.

For $r = 2$, we have
$$S_n^{(2)} = \frac{(n+2)(n+4)}{8}, \text{ if } n \text{ is even.}$$
$$= \frac{(n+1)\,(n+3)}{8}, \text{ if } n \text{ is odd.}$$
$$D_n^{(2)} = \frac{(n+1)(n+2)}{2}.$$

Hence, we obtain

$$\underset{n\to\infty}{L} \frac{S_n^{(2)}}{D_n^{(2)}} = \underset{n\to\infty}{L} \frac{(n+2)(n+4)}{8} \cdot \frac{2}{(n+1)(n+2)}$$
$$= \underset{n\to\infty}{L} \frac{(n+1)\,(n+3)}{8} \cdot \frac{2}{(n+1)\,(n+2)} = \frac{1}{4}.$$

The given series is, therefore, summable Cesàro, the sum being $\frac{1}{4}$, and the degree of indeterminancy is 2.

Borel [1] has proposed two definitions of summability, one of which employs an integral as a means of summation. His definitions are generalizations of that of Cesàro.

If we put $$c_n \equiv \frac{a^n}{n!}, \; n = 0, 1, 2, 3, \cdots,$$

where $a$ is any constant, the sequence $\{c_n\}$ decreases indefinitely as $n$ becomes infinite and the series $\Sigma c_n$ converges. Consider the function $\phi(a)$, where

$$\phi(a) = \frac{c_0S_0 + c_1S_1 + \cdots + c_nS_n + \cdots}{c_0 + c_1 + c_2 + \cdots + c_n + \cdots}. \tag{7}$$

Let us suppose the series in the numerator to converge for all values of $a$. The function $\phi(a)$ is then uniquely defined. In this discussion, the values of $\phi(a)$ take the place of the expression

$$\frac{S_0 + S_1 + \cdots + S_n}{n+1},$$

in the definition by Cesàro. If we have

$$\underset{a\to\infty}{L} \phi(a) = S,$$

then the given series is said to be summable with the sum $S$. However, we have

$$\Sigma\, c_n = e^a,$$

and by putting $$S(a) = \sum_{n=0}^{\infty} c_nS_n = \sum_{n=0}^{\infty} \frac{S_na^n}{n!}, \tag{8}$$

---

[1] See *Mémoire sur les séries divergentes, Ann. École Normale* (3), Vol. 16 (1899), pp. 9–131, 132–136; also *Leçons sur les séries divergentes*, pp. 97, 98.

this definition of Borel's can be written as follows:

$$S = \underset{a\to\infty}{L} \frac{S(a)}{e^a}. \tag{9}$$

This method of summation is not always applicable, and Borel was led to present another definition based on an integral as follows. Consider the series

$$\sum_{n=0}^{\infty} u_n = u_0 + u_1 + u_2 + \cdots + u_n + \cdots$$

and let $$S_n = u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n.$$

From (8) we have

$$S(a) = \sum_{n=0}^{\infty} \frac{S_n a^n}{n!} = S_0 + S_1 \frac{a}{1} + S_2 \frac{a^2}{2!} + \cdots + S_n \frac{a^n}{n!} + \cdots. \tag{10}$$

By differentiating $\frac{S(a)}{e^a}$ with respect to $a$, we obtain by aid of (10)

$$\begin{aligned} e^{-a}[S'(a) - S(a)] = e^{-a}\Big[(S_1 - S_0) + (S_2 - S_1)a + (S_3 - S_2)\frac{a^2}{2!} + \cdots \\ + (S_n - S_{n-1})\frac{a^{n-1}}{(n-1)!} + \cdots\Big] = e^{-a}\Big[u_1 + u_2 a + u_3 \frac{a^2}{2!} + \cdots \\ + u_n \frac{a^{n-1}}{(n-1)!} + \cdots\Big]. \end{aligned} \tag{11}$$

Putting this last series identically equal to $\phi_1(a)$, we have

$$\frac{d}{da}\{e^{-a}S(a)\} = e^{-a}\phi_1(a).$$

Remembering the definition of the sum as given by (9), we have upon integrating the above relation

$$\int_0^\infty e^{-a}\phi_1(a)da = e^{-a}S(a)\Big]_0^\infty = S - u_0. \tag{12}$$

Integrating the first member by parts, we obtain

$$S - u_0 = \left[e^{-a}\int_0^a \phi_1(a)da\right]_0^\infty + \int_0^\infty e^{-a}\left\{\int_0^a \phi_1(a)da\right\} da. \tag{13}$$

As the series defining $\phi_1(a)$ is a power series and can be integrated term by term, we have

$$\begin{aligned} \left[e^{-a}\int_0^a \phi_1(a)da\right]_0^\infty = \left[e^{-a}\left\{u_1 a + u_2\frac{a^2}{2!} + \cdots + u_n\frac{a^n}{n!} + \cdots\right\}\right]_0^\infty \\ = 0, \end{aligned}$$

since we have
$$\underset{a\to\infty}{L}\frac{a^n}{e^a}=0$$
for all values of $n$. Consequently, from (13) we obtain
$$S=u_0+\int_0^\infty e^{-a}\left[\int_0^a \phi_1(a)da\right]da. \tag{14}$$

But $u_0$ is a constant and $\int_0^\infty e^{-a}\,da=1$, and hence (14) can be written in the form
$$S=\int_0^\infty e^{-a}\left\{u_0+\int_0^a \phi_1(a)da\right\}da$$
$$=\int_0^\infty e^{-a}\left\{u_0+u_1a+u_2\frac{a^2}{2!}+\cdots+u_n\frac{a^n}{n!}+\cdots\right\}da.$$

Putting $U(a)\equiv u_0+u_1a+u_2\frac{a^2}{2!}+\cdots+u_n\frac{a^n}{n!}+\cdots,$

we have the result $\boldsymbol{S=\int_0^\infty e^{-a}U(a)da,}$

where as before $S$ is called the **sum** of the given series
$$\Sigma u_n=u_0+u_1+u_2+\cdots+u_n+\cdots.$$

As an illustration, consider the series
$$1-x+x^2-x^3+\cdots+(-1)^nx^n+\cdots$$
for values of $x>1$. We have already seen how we can associate with this series the value $\frac{1}{2}$ for $x=1$. We shall now show how Borel's method of summation enables us to associate this series with the function $\frac{1}{1+x}$ for values of $x>1$. We have
$$U(a)=1-ax+\frac{a^2x^2}{2!}-\frac{a^3x^3}{3!}+\cdots=e^{-ax},$$
$$S=\int^\infty e^{-a}U(a)da=\int_0^\infty e^{-a}e^{-ax}da=\int_0^\infty e^{-(1+x)a}da=\frac{1}{1+x};$$
that is, the given divergent series is summable with the sum $\frac{1}{1+x}$.

It can be shown [1] that these definitions of summability are consistent; that is, they give the sum in the ordinary sense when applied to convergent series. They also satisfy the boundary value conditions. In other words, if the radius of convergence of the given power series $\Sigma\, a_n x^n$ is 1 and within the interval of convergence defines the function $f(x)$, then we have

---

[1] See Ford's Studies, Arts. 37–39, *General Aspects of the Theory of Summable Series;* and also Hurwitz's article in the Bulletin of Amer. Math. Soc., Vol. 28 (1922), entitled *A Report on Topics in the Theory of Divergent Series.*

$$S = \underset{x \to 1^-}{L}\ f(x).$$

While the fact that these definitions satisfy the boundary value conditions gives some advantage in certain mathematical discussions, there are good reasons why such conditions should not be imposed as an essential part of the definition of summability.[1] These definitions likewise satisfy the laws of addition and subtraction of infinite series just as convergent series do. However, the integral method of Borel does not possess the additive property.

The foregoing definitions have been generalized by various authors and other definitions are to be found in the recent mathematical literature.[2]

It is of importance to know which, if any, of the mathematical processes which we make use of in connection with convergent series can be extended to summable series. To answer such questions, Borel[3] found it convenient to introduce what he calls **absolute summability.** He says that a given series is absolutely summable if not only the integral $\int_0^\infty e^{-a}\, u(a) da$ has a value but also each of the integrals

$$\int_0^\infty e^{-a}\, |\, u(a)\, |\, da, \int_0^\infty e^{-a}\, |\, u^\lambda(a)\, |\, da,\ \lambda = 1, 2, \cdot\ \cdot\ \cdot, n,$$

where $u^\lambda(a)$ denotes the $\lambda^{th}$ derivative of $u(a)$.

It can not be inferred that, because a series is convergent, it is necessarily absolutely summable. However, all absolutely convergent series are absolutely summable.[1]

---

[1] See Carmichael, *General Aspects of the Theory of Summable Series*, Bulletin of Amer. Math. Soc., Vol. 25 (1918), pp. 125–127.

[2] For a citation of literature and an abstract of the content of various papers pertaining to summable series, the reader is referred to Smail's *History and Synopsis of the Theory of Summable Infinite Processes*. Particular attention is called to the definitions and extensions to be found in the writings of Hölder, Le Roy, de Vallée Poussin, Knopp, and Riesz. For a more complete discussion of summable series, see Bromwich's *Infinite Series*, Chap. XI, Ford's *Studies on Divergent Series and Summability*, and Borel's *Leçons sur les séries divergentes* (1st Ed. 1901, 2d Ed. 1928). For a discussion of the more recent developments, the reader is referred to Carmichael's article in *Bulletin of the Amer. Math. Soc.* for Dec., 1918, on *General Aspects of the Theory of Summable Series* and to Hurwitz's article in the *Bulletin of Amer. Math. Soc.*, Vol. 28 (1922), entitled, *A Report on Topics in the Theory of Divergent Series.*

[3] See *Leçons sur les séries divergentes*, p. 99.

Among the properties of absolutely summable series, we have the following:[2]

(1) If the given series is absolutely summable, then a finite number of the terms can be rearranged, or a finite number of consecutive terms can be replaced by their sum without affecting either the sum or the summability of the series.

(2) If the series $\sum_{n=0}^{\infty} u_n$, $\sum_{n=0}^{\infty} v_n$, $\sum_{n=0}^{\infty} w_n$ are absolutely summable, having the sums $u$, $v$, $w$, respectively, then if we put

$$x_n = au_n + bv_n + cw_n,\ n = 1, 2, 3, \cdot\ \cdot\ \cdot,$$

the series $\sum_{n=0}^{\infty} x_n$ is absolutely summable and has the sum

$$x = au + bv + cw.$$

(3) If the series $\sum_{n=0}^{\infty} u_n$, $\sum_{n=0}^{\infty} v_n$ are absolutely summable and if we put

$$w_n = u_0v_n + u_1v_{n-1} + \cdot\ \cdot\ \cdot + u_nv_0,$$

then the series $\sum_{n=0}^{\infty} w_n$ is absolutely summable and has the sum $u \cdot v$. Hardy has shown that it is sufficient for the summability of a product that one of the factor series shall be absolutely summable.[1]

(4) If we have given a polynomial $P(u, v, w)$ having numerical coefficients and if we replace $u$, $v$, $w$ by absolutely summable series, we obtain an absolutely summable series of which the sum is equal to the numerical value obtained from $P(u, v, w)$ by replacing $u$, $v$, $w$ by the corresponding sums of the series.

Hardy has introduced the notion of **uniform summability.**[3] He based his definition on the Borel notion of summability and it is as follows:

A given series $\sum_{n=0}^{\infty} u_n(x)$ is uniformly summable in the interval $a < x < b$ if for all values of $x$ the integral $\int_0^{\infty} e^{-y}\, u(x, y)dy$ converges uniformly, where $u(x, y)$ is defined by the relation

$$u(x, y) = \sum_{n=0}^{\infty} u_n(x)\frac{y^n}{n!}. \tag{15}$$

---

[1] See Hardy, *Researches in the Theory of Divergent Series and Divergent Integrals*, pp. 22–66; also Bromwich's *Infinite Series*, p. 276.

[2] See Borel's *Leçons*, etc., p. 99 *et seq.*

[3] See *Trans. Camb. Phil. Soc.*, Vol. 19 (1903), p. 301.

On the basis of this definition it can be shown that if the terms $u_n(x)$ of the given series are continuous functions of $x$ and the series (15) converges uniformly for each value of $y$, then the sum of the given series is a continuous function of $x$ provided the given series is uniformly summable.[1]

If the series of derivatives $\sum\limits_{n=0}^{\infty} u_n'(x)$ is uniformly summable, and the series $\sum\limits_{n=0}^{\infty} u_n'(x)\,\frac{y^n}{n!}$ converges uniformly, then the series can be differentiated term by term; that is, we have [2]

$$D_x\Sigma\, u_n(x) = \Sigma\, D_x u_n(x).$$

Similarly it follows that, if the given series $\Sigma\, u_n(x)$ is uniformly summable in $(a, b)$ and the series (15) converges uniformly throughout the domain $(O, X, a, b)$, then the given series can be integrated term by term and we have [2]

$$\int_a^b \Sigma u_n(x) = \Sigma \int_a^b u_n(x)dx.$$

A somewhat different definition of uniform summability has been set up by C. N. Moore [2] in that the given series $\Sigma\, u_n(x)$ is said to be uniformly summable in the interval $(a, b)$, if for $a < x < b$ the expression

$$\frac{S_0(x) + S_1(x) + \cdot\;\cdot\;\cdot + S_n(x)}{n+1},$$

where

$$S_n(x) = u_0(x) + u_1(x) + \cdot\;\cdot\;\cdot + u_n(x),$$

converges uniformly. Upon the basis of this definition the results given above also follow.

Just as with convergent series, uniform summability is not a necessary condition for the continuity of the sum function.[3] One may, however, set up a necessary and sufficient condition for the continuity of the sum by extending to summable series the notion of **quasi-uniformity** of the approach to a limit.

---

[1] See Bromwich's *Theory of Infinite Series*, Art. 109.

[2] See *Trans. Amer. Math. Soc.*, Vol. 10 (1909), p. 400.

[3] For illustrative example and proof of the following statements, see F. J. McMackin's Columbia Dissertation (1916) on *Some Theorems of Summable Divergent Series.*

## EXERCISES

**1.** Given the series

$$1 - \frac{x}{2^2} + \frac{x^2}{3^2} - \frac{x^3}{4^2} + \frac{x^5}{5^2} - \cdot \cdot \cdot, \quad x < 1.$$

Can the terms of this series be rearranged at pleasure without affecting the convergence or the sum? Give reason for conclusion.

**2.** Given the two series $\Sigma \frac{x^n}{n!}$, $\Sigma(-1)^{n-1}\frac{x^n}{n!}$. Can they be multiplied? If so form the product.

**3.** Show that the series

$$\cos x + \frac{\cos 2x}{2^2} + \frac{\cos 3x}{3^2} + \frac{\cos 4x}{4^2} + \cdot \cdot \cdot$$

converges uniformly. Give the interval of uniform convergence. Plot some of the approximation curves.

**4.** Let $S_n = nxe^{-nx^2}$. Does it converge uniformly? Quasi-uniformly? Can the series be integrated term by term? In each case state the interval over which your conclusion holds.

**5.** Given the series $\sum_{n=0}^{\infty} (-1)^{n-1}x^n$. Show that it can be integrated term by term and thus obtain the value of log 2.

**6.** Given

$$\phi(t) = t \sin \frac{1}{t}, \quad t \neq 0,$$
$$= 0, \quad t = 0.$$

By the principle of condensation of singularities, define by aid of $\phi(t)$, a function $f(x)$ which shall be finite and continuous having no derivative at the set of rational points.

**7.** Given the series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdot \cdot \cdot,$$

show that one can by differentiation obtain the expansion of $\cos x$ and then by multiplication find the expansion of $\sin 2x$. Give reasons for each step and carry out the necessary operations.

**8.** Consider the series given in Ex. 2, Art. 57. While this series does not converge uniformly, show that it converges quasi-uniformly in the interval $(-1, 1)$. What consequences follow from this conclusion?

**9.** Show that the series

$$a^x = 1 + \frac{\log a}{1!}x + \frac{\log 2a}{2!}x^2 + \cdot \cdot \cdot + \frac{\log (n-1)\,a}{(n-1)!}x^n + \cdot \cdot \cdot$$

can be integrated and differentiated term by term. What statement can be made about the interval of convergence of the resulting series?

**10.** Discuss the existence of the integral $\int_0^\infty \frac{e^{-xy}}{1+y^2}\,dy$ for each $x$. By continued integration by parts develop an asymptotic expansion of this function for large positive values of $x$.

**11.** The function $\phi(y) = \underset{t \to 0}{L}\, y^{2t}$ has, by Ex. 2, Art. 25, the value zero for $y = 0$ and otherwise the value 1. By Cantor's method of condensation of singularities set up a function $f(x)$ having a removable singularity at the points: $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdot \cdot \cdot, \frac{1}{k}, \cdot \cdot \cdot.$

**12.** Show that the series

$$1 - 2^3 + 3^3 - 4^3 + \cdots + (-1)^n(n+1)^3 + \cdots.$$

is summable Cesàro of order four.

**13.** Given the series

$$\left(1 - \frac{1}{2^2}\right) + \left(\frac{1}{3^2} - \frac{1}{4^2}\right) + \left(\frac{1}{5^2} - \frac{1}{6^2}\right) + \cdots.$$

Without summing this series show that one may remove the parentheses without affecting the convergence or the sum of the series. Show also that the series

$$1 - \frac{1}{2^2} - \frac{1}{4^2} + \frac{1}{3^2} - \frac{1}{6^2} - \frac{1}{8^2} + \frac{1}{5^2} - \frac{1}{10^2} - \frac{1}{12^2} + \frac{1}{7^2} - \cdots$$

has the same sum as the given series.

**14.** Which of the following series converge uniformly? If uniformly convergent, state the interval within which this is true.

$$(a)\ 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots,$$

$$(b)\ \sin x + \frac{\sin 2x}{2^2} + \frac{\sin 3x}{3^2} + \cdots + \frac{\sin nx}{n^2} + \cdots,$$

$$(c)\ 1 + (x-1) + (x^2 - x) + \cdots + (x^n - x^{n-1}) + \cdots.$$

Plot enough approximation curves in each case to justify your conclusion.

**15.** In the interval (0, 1) let $f(x)$ have the value zero in all points where $x = (\frac{1}{2})^n$, $n = 0, 1, 2, 3, \cdots$, and the value 1 at all other points. Show that $f(x)$ can be represented by a convergent series each term of which is a continuous function in the given interval.

**16.** Can a series converge uniformly without converging absolutely? Quasi-uniformly and not uniformly? Can it converge without being absolutely summable by Borel's definition? Illustrate your conclusions.

**17.** State four important things concerning power series. Illustrate each with an exercise.

**18.** In Ex. 3, Art. 59, we have considered a series where $S_n(x)$, considered as a function of $n$ and $x$ together, is not bounded, although bounded in $x$ for each value of $n$. For what set of values of $x$ could this condition be true and still the series be integrated term by term over the interval containing them. If this is possible, illustrate your conclusion.

**19.** Let $f(x)$ be defined by the series

$$\frac{1}{1+x^2} + \frac{1}{8+16x^2} + \frac{1}{27+81x^2} + \cdots + \frac{1}{n^3+n^4x^2} + \cdots.$$

Show that one is justified in differentiating this series term by term in order to obtain the derivative of $f(x)$ for values of $x$ in a certain interval. Determine that interval.

# INDEX

(Numbers refer to pages.)

Absolute continuity, 102, 206, 307, 320.
Absolute convergence of infinite integrals, 240, 265, 266.
Absolute convergence of a series, 335, 340, 368, 371, 374.
Absolute summability of series, 397.
Algebraic function, 81.
Algebraic number, 17, 77.
Approximation curves, 115, 347.
Arbitrarily small, definition of, 1.
Asymptotic series, 388.
Axiom of Archimedes, 18.
Axiom of continuity, 19.

Baire's classification of functions, 144, 145, 146, 164, 324, 325.
Base of numerical system, 22.
Borel measure, 323, 324, 325.
Boundary point of a set, 30, 56.
Bounded function, 81, 215, 216, 217, 218, 219, 221, 222, 229, 242, 247, 252, 257, 260, 270, 287, 290, 291, 292, 293, 294, 315, 316.
Bounded sequence, 3, 5.
Bounded sets, 28, 61, 63, 70.

Cardinal number, 27, 48, 51, 54, 130, 138, 140, 142, 186.
Change of variables, 222, 224.
Characteristic function, 324.
Closed interval, 79.
Closed set, 30, 37, 44–46, 50, 63, 126, 128, 135.
Complementary sets, 29.
Condensation of singularities, Hankel's method, 375; Cantor's method, 385.
Conditionally convergent series, 336, 337.
Content, 58–64.
Continuity, 98–100, 107–123, 130, 135, 138, 140, 142, 161–168, 172, 173, 175, 184, 213, 216, 217, 220, 222, 224, 226–229, 251, 252, 257, 261, 267, 268, 284, 318, 352, 357, 358, 374.
Continuous functions with no derivatives, 155–160.
Continuum, 30, 39, 48, 53, 54, 56.
Convergence of infinite integrals, 240–247.
Convergent sequence, 3, 34.
Convergent series, definition of, 333.

Denjoy integral, 328.
Dense in itself, 46.
Derivative, 149, 161, 163, 164, 165, 224, 226, 227, 230, 284, 309, 312, 315, 316, 317, 319, 320, 363, 365.
Derivative number, 149, 151, 166, 167, 168, 208, 222, 224, 228, 229, 284, 309, 312, 318.
Derived sets, 28, 35, 37, 43, 46, 60.
Differentiation as inverse of integration, 230–233, 322.
Differentiation of series, 363–367, 373.
Differentiation under integral sign, 262, 263.
Discontinuity of first kind, 99, 134, 205, 214.
Discontinuity of second kind, 99.
Discontinuous functions, 127, 128, 132, 134, 135, 205, 237, 246, 283.
Discrete set, 60, 61, 63, 64, 212.
Divergence of infinite integrals, 240, 242.
Divergent series, 333, 387–399.
Double integrals (Riemann), 268–274; (Lebesgue), 289, 325–326.

Enumerable set, 28, 37–43, 44, 47, 50, 51, 52, 55, 63, 68, 72, 73, 134, 206, 257.
Equi-absolute continuity, 301, 302, 305, 363, 366.
Equivalent sets, 31.
Everywhere dense sets, 28, 47, 52, 135.
Exterior measure, 65, 67, 68, 69, 70.
Exterior point of a set, 30.

Finite integral, 199.
First category, sets of, 50–55.
First species, sets of, 36, 43, 61.
Fundamental theorem of integral calculus, 229, 322.

Heine-Borel theorem, 44, 45, 57.
Hellinger integral, 328.
Higher derivatives, 152, 153.

Indefinite integral (Riemann), 225, 226–233; (Lebesgue), 306–322.
Infinite integrals, 199, 234–250, 297–306, 322.
Inner content, 59.
Inner limiting set, 132.
Inner point of a set, 29.
Integrals involving a parameter, 250–264.
Integration of series, 358–363, 373.
Interior measure, 65.
Interval, 21, 41.
Interval of convergence of a series, 334.
Irrational numbers, 7, 41; Cantor's theory of, 9; Dedekind's theory of, 11.
Irrational points, 20, 41.
Isolated point, 29, 46.
Isolated set, 29, 41.
Iterated integrals, 269, 270.
Iterated limit, 85.
Interchange of order of differentiation, 189, 190, 192, 194.
Interchange of order of integration (Riemann), 270; (Lebesgue), 325.

Jordan measure, 59.

Law of the mean for differential calculus, 160; for integral calculus, 220–222, 290.
Lebesgue integral, definition of, 285; condition for existence of, 287; comparison with Riemann integral, 295, 321; for non-bounded functions, 297–306; of two independent variables, 323–327.
Left-hand derivative, 149.
Left-hand limit, 85.
Limited variation, 202, 203, 204, 205, 206, 208, 214, 308, 319.
Limit of a sequence, 33; of a function, 82; condition for the existence of, 95, 97.
Limits, laws of operation with, 88–97.
Limiting function of a sequence, 82, 281, 282, 294, 295, 301, 302, 304, 305.
Limiting point of a set, 28, 35.
Linear continuum, 39.
Linear set defined, 25.
Lower boundary, 31, 81, 219.
Lower boundary function, 82, 281.
Lower integral, 200, 223, 269.
Lower limit, 31, 81.
Lower limiting function, 82, 282, 284.
Lower semi-continuous function, 123, 124, 126.

Maximum, 81, 109.
Measure, 64, 70, 72–76, 212, 215, 283, 299.
Measurable function, 277–284, 287, 290–295, 315, 362.
Monotone function, 214, 215, 221, 222, 224, 242, 247, 317, 318.
Monotone sequence, 4, 5, 281.
Multiple-valued function, 80.
Multiplication of series, 342–345, 374.

Natural numbers, 1.
Neighborhood, 21.
Non-enumerable sets, 28, 39, 41, 42, 46, 50.
Non-linear sets, 56.

Open set, 30, 44–46.
Ordered set, 30.

Ordinary discontinuity, 99, 205, 214.
Oscillation, 100, 101.

Parentheses, removal of, in a series, 338, 339.
Partial derivatives, 170, 172, 173, 175, 185, 186, 187, 190, 192, 194.
Perfect set, definition of, 30; properties of, 46–50.
Perron integral, 330.
Point of condensation, 29, 42, 44, 46.
Point-wise continuous function, 132, 138, 140.
Point-wise discontinuity, 104, 127, 129, 130, 175, 353.
Power series, 367–375.
Primitive function, 227.
Principal part of infinite integral, 235.
Product of point sets, 32.

Quasi-uniform convergence of functions, 121, 122, 257; of infinite integrals, 265, 267; of infinite series, 355–358, 359, 360, 362.
Quasi-uniform convergence in general of functions, 122, 256; of infinite integrals, 265, 268.

Radius of convergence of power series, 369.
Rational function, 81.
Rational integral function, 81.
Rational numbers, definition of, 1; sequence of, 3; partition of, 11.
Rational points, 18, 39, 51.
Real numbers, definition of, 8; system of, 12, 13, 14; sequence of, 14; geometric representation of, 17.
Reducible set of points, 237.
Regular sequence, 4, 5, 15.
Removable discontinuity, 99.
Riemann integral, 198, 209–221, 224, 229, 237, 251, 252, 256, 257, 260, 295.
Right-hand derivative, 149.
Right-hand limit, 85.

Saltus, 101.
Second category, sets of, 50, 53, 54.
Second species, sets of, 37, 43.
Semi-continuous functions, 123–127, 128.
Sequence defined, 3; regular sequence, 4, 5; bounded sequence, 3, 4, 5; convergent, 3; of intervals, 22.
Sequence of point sets, 33, 75, 76, 290, 294; complete limit of, 33; restricted limit of, 34; convergence of, 3, 34.
Simply-uniform convergence of series, 353–355.
Simultaneous limit, 86, 91.
Single-valued function, 80.
Stieltjes integral, 327.
Subtraction of point sets, 32.
Subtraction of series, 342.
Sum of point sets, 32, 37.
Sum of series, 342.
Summable function, 298–308, 312, 318, 320, 362, 363.
Summability of series, 391; Hölder's method, 392; Cesàro's method, 393; Borel's method, 394.

Total differentiability, 177, 180–187.
Total variation, 202, 308.
Transcendental function, 81.
Transcendental numbers, 17.
Transfinite numbers, 26–27, 43.

Uniform continuity, 103, 107.
Uniform convergence of functions, 115–119, 142, 164, 187, 260; of integrals, 264–268; of series, 345–353, 361, 363, 371.
Uniform summability of series, 398.
Upper bound, 31, 219.
Upper boundary function, 82, 281.
Upper integral, 200, 223, 269.
Upper limit of a set, 31; of a function, 81.
Upper limiting function, 82, 282, 284.
Upper semi-continuous function, 123, 125, 126, 174.

Variation, see limited variation.

Weierstrass test for uniform convergence of series, 349.
Well-ordered set, 31, 37.

Date Due

AR 2 4 1959
OC 3 0 '63
NO 20 '63
FE 12 '64

QA331
.5
.T74

30548

ST. OLAF COLLEGE
QA331.5 .T74
Townsend, Edgar - Functions of real variables...
3 0111 00357 3465